Proceedings of the Twenty First Annual Meeting of the International Society for Psychophysics

Traverse City, Michigan, USA
October 19-22, 2005

Fechner Day 2005

Editors

John S. Monahan, Sonya M. Sheffert Central Michigan University
James T. Townsend Indiana University

Reference

J.S. Monahan, S.M. Sheffert, & J.T. Townsend, (Eds.). (2005). *Fechner Day 2005. Proceedings of the Twenty-first Annual Meeting of the International Society for Psychophysics.* Mt. Pleasant, MI: The International Society for Psychophysics.

Front Cover Illustration: Old Mission Peninsula courtesy of the Traverse City Convention and Visitor Bureau.

Spine Illustration: Michigan and Traverse City courtesy of Northern Michigan College.

Distribution:

International Society for Psychophysics
c/o John Monahan
Department of Psychology
Central Michigan University
Mt. Pleasant, MI 48859

Telephone 989-774-6491
Fax 989-774-2553
Email: monah1js@cmich.edu

Printed in USA by:

Central Michigan University Printing Services
Mt. Pleasant, MI USA 48859

INTERNATIONAL SOCIETY FOR PSYCHOPHYSICS
Founded in Cassis, France, October 22, 1985

Executive Committee 2004-2005

Preface

Welcome to Traverse City, the Cherry Capital of Michigan. We hope you enjoy your stay in this lovely resort town and get a chance to see the beautiful autumn colors.

The impact of Fechner's research approach extends far beyond psychophysics; many different disciplines and intellectual pioneers, from Wundt to Freud, were influenced by Fechner's pioneering research and methodology. Similarly, the presentations at Fechner Day 2005 represent a wide range of perspectives and applications and span the continuum from theoretical to applied. As Lewin said, "Nothing is as practical as a good theory," and our papers reflect the accuracy of that hypothesis with theory applied to both academic and non-academic problems.

We thank Mark Elliott, whose helpful and timely advice made the conference possible in the compressed time frame allowed. Thanks to Steve Link, Ron Verillo, and George Gescheider for preparing tributes to our fallen comrades, Doug Vickers and Sandy Bolanowski. We are grateful to David Rosenbaum for agreeing to be the keynote speaker. We also thank the symposia organizers: Ken Norwich and Willy Wong for "Putting 'Physics' back into 'Psychophysics"; Ehtibar Zdafarov for "Measurement in Perception"; Simon Grondin for "Processing Temporal Information"; and Åke Hellstrom for "Stimulus Comparison Process." And thank you to all who are presenting spoken and poster papers.

We owe a special debt to The Department of Psychology of Central Michigan University for providing administrative assistance and funding. The strong support we received from the former chair, Gary Dunbar, and current chair, Hajime Otani, has been invaluable. Dean Gary Shapiro, Provost Tom Storch (both from CMU), and Scott Parker also provided funds to support graduate travel. The Office of Research and Sponsored Programs furnished the poster boards and easels. Lastly, we thank Barbara Houghton, Kay Purtil, and Joan Maul for administrative support.

John S. Monahan
Sonya M. Sheffert
James T. Townsend

CONTENTS

Tributes to Fallen Colleagues

Papers in Alphabetical Order

DOUGLAS VICKERS
(1940 - 2004)
(Photo courtesy of Dr. Ted Nettlebeck)

IN MEMORIAM

Douglas Vickers
(1940 – 2004)

Stephen Link
University of California, San Diego

Douglas Vickers, age 64, died suddenly at his home in Adelaide, Australia, on October 31, 2004. He is survived by his wife Yvonne, children Marc and Anne, and six grandchildren. His astonishing career as a theoretical psychologist, experimental scientist, and mentor to famous students spanned forty years. Held in the greatest esteem, his students refer to him as a teacher of unfailing sincerity who was generous and warm in his dealings with others.

Vickers' education began in the Scottish city of Dunbar. He advanced with distinction, graduating from Edinburgh University in 1961, and continued to an Honours BA Degree from Cambridge University where he placed first in of the Natural Sciences Tripos (Part 2). In 1967 he received his Ph D supervised by Alan Welford. Later, in 1994 Cambridge honored him with its Sc.D. In 2000-2002 Douglas Vickers served as President of the International Society for Psychophysics. Interspersed within this breathtaking career are the major undertakings that brought him international fame.

Douglas' scientific career began in earnest with the famous Shallice and Vickers (1964) paper "Theories and experiments on discrimination time" appearing in volume 7 of Ergonomics. This important review of psychophysical models provided an English approach to a topic of deepening interest in Europe and the United States. His PhD thesis "Visual discrimination and the perception of visual depth" (1967) continued his exploration of perceptual phenomena investigated by psychophysical means. But the lure of Australia caught his interest at the time when Australian universities, expanding to meet the needs of a post WWII society, sought outstanding talent from around the world. Professor Malcomb Jeeves, Head of the Psychology Department at Adelaide University, offered Douglas a position. In 1967 Doug immigrated to Australia to become Lecturer in Psychology.

At the time Alan Welford championed investigations of models of the time taken to perform judgment tasks. The application of Abraham Wald's Sequential Probability Ratio test to human choice reaction experiments investigated by Stone at the Applied Psychology Research Unit, Cambridge, in 1960 did much to inspire an English approach to this topic. By 1970 several approaches vied to provide accounts of response time data gathered from both discrimination and multiple choice experiments. Among them were the finite state models proposed and investigated by Ollman, Yellott, and Link, the stochastic models of La Berge, Audley, Christie and Luce, McGill, Hohle, and Link, and the sequential theories of Stone, Edwards, and Laming. Then, in 1970, Vickers added new, compelling, ideas to the ongoing scientific adventure in the fundamental paper "Evidence for an accumulator model of psychophysical discrimination."

This paper begins a lifelong development and application of an elegant approach to modeling comparative judgments. When discriminating between two stimuli there are three classic measures of performance, the proportion of correct responses, the length of time taken to discriminate, and the subjective confidence that the discrimination made is correct. Presumably a single discriminative mechanism is the basis for each of these responses. Vickers idea was that this mechanism consisted of counters associated with each possible response. The counter that reached a pre-set threshold first gave rise to a response. An illustration of the structure of the model and an illustrative stochastic path appear in Figure 1.

The original theory (Vickers, 1970) can be imagined as a stochastic development of Fechner's theory of sensory representation and Thurstone's idea of comparative difference. Two physical stimuli, A and B, are transformed into internal normally distributed random variables s_A and s_B (the Fechnerian theory of sensory representation). A Thurstonian assumption gives rise to comparative differences between two stimuli. The result is a normally distributed distribution of sensory differences for $s_A - s_B$. Suppose A > B so that the mean of $s_A - s_B$ is positive. It is more likely that a value of the difference $s_A - s_B$ will be positive. Call this probability p. Of course there will then be a probability $q = 1 - p$ that a value of $s_A - s_B$ will be negative.

The accumulator model posits that the time of making a decision is filled by sampling values of $s_A - s_B$ and adding positive values to a counter T_A and negative values of $s_A - s_B$ into a counter T_B. If the counter T_A reaches a maximum (threshold) value of C_A before the T_B counter reaches a maximum of C_B then the response R_A occurs to indicate that stimulus A is greater than stimulus B. Otherwise, the negative counter T_B will exceed its maximum (threshold) and response R_B occurs indicating that stimulus B is smaller than A. The threshold for responding values, C_A and C_B are under control of the subject and proposed to be related to the variability of the sensory representations, s_A and s_B.

The simplicity of the theory is its charm: two stimuli, a single distribution of sensory differences (á la Thurstone), two accumulators, and two response outcomes. The beauty of the theory is the vast number of predictions made about response probability and response time. In a way the two counters race to be the first to generate a response. Sometimes, due to the variability of sensory differences s_A and s_B, the negative counter will reach its criterion C_B before the positive counter reaches its criterion, C_A. In this case the response R_B will occur in error and stimulus B will be judged greater than A.

A stochastic path illustrating the outcome for a single trial for which the distribution of sensory differences is N(0.2, 1.0) appears in Figure 1. The Y axis represents the amount of positive difference values accumulated during a trial. The X axis represents the amount of negative difference values accumulated during the trial. The use of "amount" allows the negative values to be represented as positive increments to an ongoing total. Two response thresholds are shown with values of $C_A = 10$ and $C_B = 12$. That is, more negative values are required to response B>A than are needed to respond A>B. The difference between these response thresholds characterizes a bias in favor of response R_A.

VICKERS' (1970)
ACCUMULATOR THEORY OF DISCRIMINATION

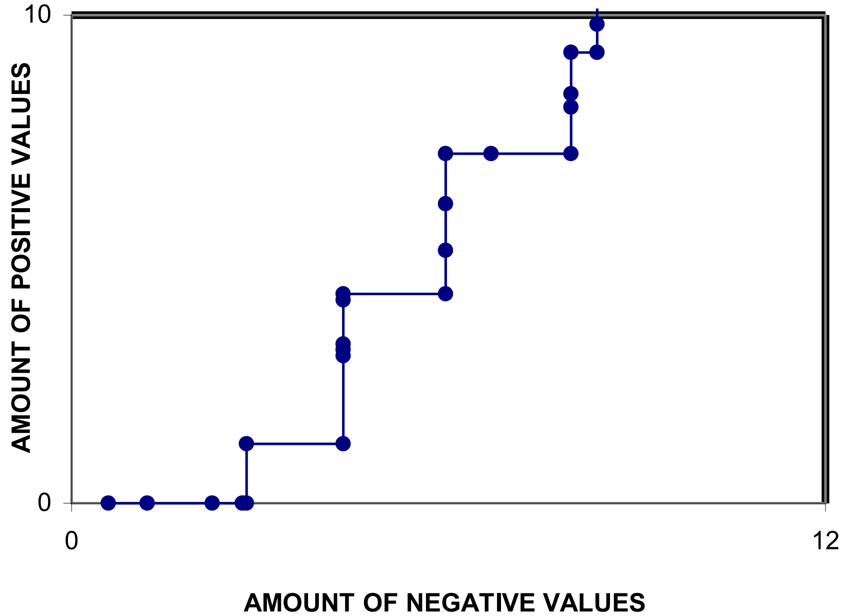


Figure 1. A single stochastic path illustrating the accumulation of positive and negative comparative differences. In this case five negative differences occur by chance and generate the first increments along the lower boundary. Then a positive difference occurs followed by another negative difference. The accumulation continues until a boundary is breached, in this case the upper bound.

A variable number of differences s_A - s_B will occur before a response criterion is reached. This duration is the model's predicted decision time. In Figure 1 the particular sequence of observations creating this single within-trial accumulation of difference values leads to response R_A in a total of 24 observations. The last difference, which pushes the positive values beyond the threshold at ten, is not shown. If each of these sensory differences requires an average amount of time Δt the then total decision time for response R_A is $24\Delta t$. The response R_A is a correct response because stimulus A is greater than B.

At the time theoretical interest focused especially on the analysis of correct and error response times. The importance of this result stemmed in part from the proof by Stone that for the Sequential Probability Ratio Test, invented by Abraham Wald, these responses must be identical with respect to decision time. Laming (1968) investigated this prediction and found it to be false. Simulations of the Vickers' Accumulator Model showed differences between correct and error decision times to be a characteristic theoretical prediction that distinguished it from the previous random walk models that required correct and error decision times to be identical.

The 1976 meeting of Attention and Performance VII in Sénanque, France, offered Douglas an opportunity for an important advancement in the accumulator theory. The title of the paper, "An adaptive model for simple judgment," does not prepare one for the major extensions suggested and empirically evaluated. The two accumulator model is described and then a three category model is described to account for, in part, judgments of sameness and difference. Confidence in simple judgments is addressed and the manner in which parameters of the theory adapt to experimental conditions discussed. On a personal note, we both attended this meeting. It was my first opportunity to discuss directly with Doug the predictions of the accumulator theory and the distribution-free random walk model I proposed in 1975. I was most impressed by Doug's knowledge of the literature and his ideas about a general theory of psychophysics. And, he was already a likable Australian.

Then, in a stunning contribution to psychophysical theory Vickers' provided further theoretical investigations and elaborations that accounted for the third important measurement variable, response confidence. In "Decision Processes in Visual Perception" (1979) Vickers provided a broad survey and theoretical investigation of simple decision processes, confidence and adaptation, and complex decision processes. This monograph followed up Vickers' previous paper (Where Angell feared to tread: response time and frequency in three-category discrimination, 1975) by proposing a theory of equal judgments that yielded a convincing account of much experimental data. With references to 538 important experimental and theoretical papers, this masterful treatment of decision processes merits the attention of modern scholars, and is one of the most important psychophysical publications in the 20th century.

Decades of investigations into the underlying mechanisms of human perception and cognition followed this propitious beginning. The following jointly-authored articles provide an impression of the many different ideas and applications of Douglas' contributions: "Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task" in 1982, "The accumulator model of two-choice discrimination in 1988, "Interoperative patient controlled sedation" in 1991, "Intelligence and visual and auditory discrimination" (1995), "Towards a dynamic constructionist model of memory," (1996) "Memory capacity and intelligence," (1997), "Never cross the path of a traveling salesman," (1998), "Human performance on visually presented traveling salesman problems," (2001), "Towards a generative transformational approach to visual perception," (2001), "Decision making and memory," (2001), and "The perception of minimal structures," in 2003 - just a sampling of his wide-ranging scientific interests.

Douglas gave members of the International Society for Psychophysics (ISP) new ideas on a regular basis. His publications at the ISP meetings began in 1989 with "Stopping rules in discrimination models," and continued with "Individual differences in the rate of accrual of information, the FAST (frequency accrual speed test) index," (1990), "Confidence and memory in simple judgments," (1999), "Decision making and memory: Predicting accuracy, response time and confidence on individual trials," (2000), "Where does the balance of evidence lie with respect to confidence?" (2001) and "Confidence and time in three - category judgments" (2003). Each a delightful contribution.

His contributions to international scientific activities were extensive and welcomed by his colleagues. He was the organizer of many scientific sessions including “Human Decision and Choice” at the Twelfth Annual Conference of the Australian Psychological Society (1977), “Decision and Control Mechanisms in Perception and Memory” at the XXIV International Congress of Psychology (1988), “Theoretical Modeling in Psychophysics ” at the Sixteenth Annual Meeting of the International Society for Psychophysics (2000), and “Confidence and Psychophysical Judgments” at the Nineteenth Annual Meeting of the International Society for Psychophysics (2003). He was a member of the Organizing Committee of the 29th Experimental Psychology Conference in Adelaide, Australia, (2002). His contribution to us was to be President of International Society for Psychophysics (2000-2002) and Organizer for the Twenty-First Annual Meeting of the International Society for Psychophysics to be held in Adelaide, Australia in 2005.

Beyond these major events Doug also provided the field with his capabilities as reviewer of journal submissions, grant applications, manuscripts, conference submissions and professional appointments. He was a founding member of *Non-linear dynamics, Psychology and Life Sciences* and Editor, with P. L. Smith, of *Proceedings of the XXIV International Congress of Psychology* (Volume 2) entitled *Human Information Processing: Measures, Mechanisms and Models*. He reviewed articles for such major journals as *Psychological Review, Journal of Mathematical Psychology, Acta Psychologica, Perception, Journal of Experimental Psychology: Human Perception and Performance, Psychonomic Bulletin & Review, Perception & Psychophysics, Memory & Cognition*, *Australian Journal of Psychology, Perceptual and Motor Skills, Personality and Individual Differences,* and *Nonlinear Dynamics, Psychology and Life Sciences.*

His memberships in scientific societies expressed his vigourous interest in advancing our scientific knowledge of mental processes. He was, for example, an Advisory Council Member of the International Association for the Study of Attention and Performance (1978-83), a member or associate of some twelve international learned societies, and President of the International Society for Psychophysics (2000-2002).

But, this marvelous career came to an unexpected halt on Sunday, October 31, 2004. Now, a voice is missing from among us. Our friend, colleague, mentor, our excellent scientist is quiet. The silence is deepened by our remembrance that this year we were to be his guests in Adelaide, Australia. Yet, I believe that in appreciation of this remarkable life, we should celebrate his wonderful career - a career peopled by famous teachers, outstanding contributions to our science, memorable lectures, insight, wit, intelligence, esteem, and international recognition. We can celebrate his many contributions through our fond memories of a charming colleague, an excellent scientist, and, to so many, a very good friend.

Tribute to Sandy Bolanowski (1950-2005)

Ronald T. Verrillo and George A. Gescheider
Syracuse University and Hamilton College

It is a privilege and honor to be here to present this tribute to Sandy Bolanowski. Sandy died suddenly and unexpectedly on January 13, 2005 while on leave at Vanderbilt University in Nashville, Tenn., where he had gone to perform a series of physiological studies on the somatosensory system with John Kass. This meeting of our society was to be hosted by Sandy in Syracuse following the untimely death of Douglas Vickers who was to host this meeting originally scheduled in Australia. We salute the bravery of John Monahan for assuming the risk of hosting the meeting here in Michigan. Keep your fingers crossed, John, the meeting isn't over yet.

Sandy was born Stanley J. Bolanowski on February 22, 1950 in Utica, New York, where he attended both grammar and high school. He originally chose the field of electrical technology and earned an Associate Degree in 1970 at the State University of New York at Morrisville, NY. He continued his studies at Syracuse University where he earned a bachelor's degree in psychology in 1973 and a bachelor's degree in electrical engineering/biosystems in 1974.

It was during his undergraduate studies at Syracuse that we first met Sandy while he was working with Ron Verrillo, who founded the somatosensory laboratories at the Institute for Sensory Research. In 1981 Sandy received a doctorate in Sensory Sciences. He had built the first somatosensory physiology laboratory at ISR and utilized it for his PhD dissertation entitled “Intensity Frequency Characteristics of Pacinian Corpuscles.” Dr’s Ron Verrillo and Jozef Zwislocki were his graduate advisors.

He was awarded a post-doctoral fellowship at the Center for Brain Research at the University of Rochester School of Medicine where he was appointed assistant professor and assistant professor at the Center for Visual Science at the University of Rochester. He held both of these appointments from 1983 to 1987. During his tenure there he studied the cortical activity within the visual system of monkeys as well as the psychophysical responses of humans to *ganzfeld* stimulation. In this research he collaborated with the world-renowned neurophysiologist Dr. Robert Doty.

In 1988 Dr. Bolanwoski returned to Syracuse University where he continued his collaboration with Ron Verrillo and George Gescheider of nearby Hamilton College, Department of Psychology. It should be noted that throughout his tenure in Rochester Sandy continued to work on somatosensation in Syracuse, making periodic trips between the two cities. His dedication to sensory science knew no bounds of interest and energy. In 1988 he published a paper in the Journal of the Acoustical Society of America entitled “Four channels mediate the mechanical aspects of touch” with co-authors, George Gescheider, Ron Verrillo and Christine Checkosky. This landmark paper has been described as the “now classic paper that provided evidence of four different mechanoreceptors…” with basic findings that to this day “remain the best theory of mechanoreceptor function” (J.E. Weisenberger, 2004 “History of Bioresponse to Vibration in the Acoustical Society of America, ASA at 75” p. 81.)

Appointed associate professor of bioengineering and neuroscience in 1988, he was promoted to full professor in 1994, a position that he held until the time of his death. He was director of the Computational Neuroscience Program from 1995 to 1999 and in 2002 he was appointed Associate Director of the Institute for Sensory Research under its current director Dr. Robert L. Smith. After his return to Syracuse in 1988 he continued his association with the University of Rochester where he was adjunct professor in the Department of Surgery from 1988 to 1992. He was a visiting professor in the Department of Otolaryngology at the Keck Center for Integrative Neurobiology and the Coleman Laboratory at the University of California at San Francisco from 1992 to 1993. In California he collaborated with Dr. Michael Merzenick on the structure and function of the sensory systems of pain, temperature and touch. He collaborated often with scientists outside of his home base at Syracuse. This activity included Dr. M. Holmes, Department of Mathematical Sciences at Renssalear Polytechnic Institute, Dr. J. Bell, Department of Mathematics at SUNY, Buffalo. With these colleagues he modeled the mechanics and transduction mechanisms of the Pacinian corpuscle. With Dr. D. Dishman at the NY Chiropractic College, Seneca Falls, NY he studied the double-crush syndrome and other nervous system disorders and developed a manipulator for the chiropractor to control spinal manipulations. With Dr. A. Apkarian at the SUNY, Health Science Center at Syracuse, he did research on pain-touch interactions, functional MRI and stochastic resonance. In conjunction with U.S. Navy scientists and Ron Verrillo at the Naval submarine Medical Research Laboratory in Gaston, Connecticut he designed psychophysical and clinical tests to evaluate the sensory capacity of deep-sea divers. He also participated in the design and execution of experiments to evaluate devices to aid in the situational awareness of pilots at the Naval Air Medical Research Laboratory at Pensacola, Florida.

Sandy received many research grants and contracts from government, military and industrial institutions. In addition to his academic work, he founded and was CEO of his own research company, based in Skaneateles, NY. His expertise spanned many areas of sensory science including neurophysiology, anatomy, psychophysics and the modeling of sensory systems as well as the engineering skills to develop and test practical devices for use in laboratory, industry and the clinic. We would list among his contributions to sensory science: the precise determination of the intensity and frequency characteristics of the Pacinian corpuscle at the levels of both the receptor and spike potentials, the effects of response criteria and skin temperature on the response to vibration of humans and in excised Pacinian corpuscles, the effects of monocular and binocular contourless stimuli on the visual responses of humans and in the cortical activity of monkey's, the modeling of physiological and psychophysical responses of the tactile sensory channels in human glabrous and hairy skin, the determination of the columnar compression patterns in the lamellae of Pacinian corpuscles by direct video observation, and the distribution patterns of rapidly and slowly adapting mechanoreceptors in cutaneous tissues. His engineering skills were manifested in the numerous laboratory instruments that he developed and perfected. Some of these systems are used today in research laboratories around the world.

An active member of a number of scientific societies, he belonged to 12 societies including the Acoustical society of America where he became a Fellow in 1993, The Psychonomic Society, the Society for Neuroscience, the Association for Research in Vision and Ophthalmology, and of course the International Society for Psychophysics.

As it is with all of us who work within academic institutions, Sandy spent a considerable amount of his time in teaching, supervising and advising. He taught classes, seminars and laboratories nearly every semester, both at Syracuse University and by invitation at nearby institutions. He taught and advised students at the undergraduate, masters and at doctoral levels; and naturally he supervised doctoral students in their dissertation research and publications. In the area of publications, he had published well over 100 papers in peer-reviewed scientific journals and books. We should not overlook the numerous oral presentations and invited addresses he delivered at meetings of scientific societies and at other academic institutions.

But, enough of the many attributes that Sandy had in the workplace, in the professional context and of the significant contributions he made to his science. We certainly join in our admiration of his creativity and his dauntless pursuit of knowledge. We should also like to share here some of what he conveyed as a friend and colleague in and out of the laboratory. We shared many, many values about science, about society and about the enjoyment of life.

When we first met Sandy he was an undergraduate and, typical of that stage of life, he was faced with many doubts and uncertainties. He was open and honest and shared his concerns with us; he was able to talk them out and came to resolve these problems successfully. We appreciated his trust and honesty and it was gratifying to see him put a troublesome phase of his life behind him.

We all shared a love of food and wine. He loved to cook and he had an impressive knowledge of the wines of the world that was accompanied by an excellent personal wine cellar. We still have a vivid memory of his putting the final touches on a baked Alaska with a blowtorch! Hardly a month went by that we did not exchange information and a critique of a new restaurant that we had discovered, which led to many culinary adventures.

As a colleague in the workplace, he was sheer pleasure. We all had a deep respect for each other which showed itself in how quickly we could agree on a course of action in the research, what needed to be done and the order in which to do them, the best way to solve a particular problem and what the data meant once it was gathered. We worked hand-in-glove and we opened each other's eyes to new meanings and frontiers. How else could we have functioned as a productive research team for over 30 years?

There was one exception to the almost perfect mesh of personalities that we had: Sandy and Verrillo were inveterate complainers, Gescheider preferred the smoother path. With Verrillo, Sandy complained about practically everything politics, the country, the university (a favorite), and of course the world. When Verrillo's wife Violet once pointed out jokingly that he was complaining again, Sandy joked in return; "If you don't complain, what is there to talk about?" To Sandy this was all in fun and we all certainly found so much in life that required no complaints and so much that was there for the sheer joy of it.

Sandy, we shared success and disappointment, ups and downs, joys and sorrows, but in the final analysis we made our contribution and enjoyed doing it to the utmost. Be assured that most deeply and sincerely, you will be missed.

VAN RIPER MEETS FECHNER: SCALING COGNITIVE AND SPEECH MOTOR SKILLS FOR FLUENT SPEECH

Joseph G. Agnello, Ph.D., Emeritus, Colleen Kato, Melissa Pogir, M.A., CF-SLP, Zoi Vavva, M.A.
University of Cincinnati, Cleveland State University, Health South at Austin, TX, Cleveland State University, Cleveland, OH
E-mail: agnello@uc.edu

Abstract

Extensive scaling procedures have been incorporated into the Van Riper therapy approach. Clients and therapists independently scale cognitive features (e.g. attitudes) and motor features (e.g. speech initiation, prosody) during the process of intensive group, situational and individual therapies. The data are based on numerous groups over a period of 20 years. A comprehensive therapy model will be presented.

This paper focuses on the constructs between the Van Riper therapy approach for stuttering and Gustav Theodor Fechner's psychometric measurements. The Van Riper stuttering model is briefly explained. Followed by a brief description of Fechner's psychometric measurements. Our current approach to fluent speech using scaling techniques is discussed. Using the therapy model that incorporates aspects from Van Riper and Fechner, data from scales that were completed by individual client's pre and post therapy are presented.

Charles Van Riper's Approach to Stuttering Therapy

The Van Riper approach to stuttering therapy follows a sequence of phases, which overlap one another and are maintained throughout the client's life. The therapeutic sequence includes four phases: Identification, desensitization, modification, and stabilization (Van Riper, 1973). As the client moves through the phases, the focus of therapy is "on the present time", or what is translated from the German term, "Conscious Present." This focus on the "Present" moment of talking is necessary in order for the client to generalize the learned fluency techniques outside of the therapy setting. In the following paragraphs each phase is explained.

The Identification Phase

During the identification phase of therapy, the therapist identifies the types of stuttering behaviors that occur, and the client identifies and classifies their own stuttering behavior, as well as their feelings and emotions behind the stuttering. This provides information about how aware the client is of his stuttering. As the client identifies his behavior, feelings and emotions, other approaches of therapy are introduced: the desensitization.

The Desensitization Phase

The purpose of the desensitization is to decrease the speech anxieties and other negative emotions that the client experiences. A well-documented way to achieve this is by pseudostuttering. Pseudostuttering takes place in a controlled environment (i.e. therapy room). During pseudostuttering exercises real stuttering will occur. Through the "realness" of pseudostuttering, clients learn to control their fluency and learn that the majority of people will not penalize them for stuttering. It is essential for clients to understand that they are in control of what they are doing and what they can do. This role playing while pseudostuttering may result into stuttering that takes place in real life situations. More important, the purpose of pseudostuttering is for the client to recognize and control the feeling of helplessness or out of control. Another feeling that is targeted with pseudostuttering is a sense of what speech aspects are involved in fluent speech by focusing on the occurrences of fluent speech. Options are of no leering benefit if one doesn't feel the sense of self-control.

The Modification Phase

The modification phase focuses on decreasing stuttering and helping the client to unlearn the avoidance and struggle responses he uses. The therapist teaches various fluency controls (i.e. smooth speech initiation or easy onset) and rates the client on how well he uses the techniques in different situations that progressively become more difficult. The therapist may use computer programs in therapy to monitor the client's use of fluency techniques. The ultimate goal in this phase is for the client to learn to monitor and rate himself independently. At this point, the stabilization phase will begin.

The Stabilization Phase

As soon as the client learns to use all of the fluency-enhancement strategies and has the appropriate attitude towards his stuttering, the next step is for him to become more independently manage his own fluency. During the stabilization the client begins to monitor his own gains of fluency and generalize his newly learned behaviors to different environments. At this level of therapy, the clinician's role is that of a consultant. The client will reconsider his views about himself and will identify the changes that he has noticed. More importantly, the client is responsible for identifying and modifying the stuttering occurrences. This is accomplished while the client pseudostutters in the therapy setting. This will help the client to use the learned techniques in real-life situations. During this phase, the client's perception of his own normal speech patterns is increased. At some point in this stage the clinician and the client decide together when to terminate therapy.

It is important to reinforce that these phases do not stand-alone entirely and do not have a timeline. Therapy is approached in the phases mentioned, but is individualized to meet the needs of the client.

A Synopsis of Gustav Theodor Fechner's Contribution

The work of Gustav Theodor Fechner (1801-1887) signified the formal beginning of experimental psychology. Fechner's research laid out the future application for the psychophysics, the unity of mind and body. Psychophysics illustrates that an increase in bodily energy corresponds to an increase in mental intensity (Wozniak, 1995).

Van Riper's Approach to Fluent Speech using Fechner's Theory

This section presents the characteristics of stuttered speech followed by a discussion about the relationship between Van Riper's approach to stuttering therapy and Fechner's theory.

A stutter event is defined by Wingate (2001) as the inability to proceed in the speech sequence and is defined by the following motor characteristics:

1. Disruptions in the *flow* (fluency) of verbal expression
2. Involuntary audible or silent *repetitions* and/or *prolongations* in short speech elements (sounds, syllables, words of one syllable)
3. *Frequent* breaks in the flow of speech
4. *Accessory movements* (motor or phonic tics) may or may not be present (e.g., fast eye-blinking)

The hallmarks of stuttering are the involuntary actions and frequent breaks in the flow of speech production characterized by repetitions of initial phonemes (Wingate, 2002). Wingate explains that these repetitions of initial phonemes may contain a vowel-like element, the schwa sound (e.g., ba-ba-ba-book). According to Van Riper (1982), stuttered speech contains more than two repetitions per word or per 100 words. Additionally, in stuttered speech, tension is often present in the vocal tract. Tension may result in interruptions of airflow, phonatory arrest (during which no sound is produced) and restricted pitch range (resulting in a monotonous voice). The more physical or muscular effort speech takes, the more breaks, blocks, and glottal fry occur in the speech output.

It is obvious that stuttering presents as a motoric event in an individual's speech. According to Van Riper (1971), in situations with high stress, a person who stutters is more likely to experience more pronounced or frequent stuttering. This suggests that the individual's thoughts or feelings are likely to affect his speech behavior.

Fechner's central idea is that the mind (or mental phenomena) and body (or matter) relate with each other in a dynamic way, in that one influences the other. The dynamic connection between mind and body is evident in several ways within the Van Riper's constructs. Agnello refers to this dynamic connection as bringing stuttering therapy to the "Conscious Present," in order for the speaker to have better control of speech in "real life" situations.

As mentioned earlier, there are four constructs of stuttering therapy: identification, desensitization, modification, and stabilization. Although stuttering therapy should be individualized to a client's needs, it should contain the following components listed in Table 1. The basis for this therapy approach is not what caused the motoric breakdown, but rather, what is happening mentally and physically before, during, and after a stuttering event has occurred. The approach is based on the core idea that stuttering increases when a client avoids or struggles on feared words, sounds or during certain situations, which cause anxiety or even fear.

Table 1, illustrates that cognition (the client's thoughts and feelings) and motor behavior are addressed simultaneously throughout therapy. Thus, the importance of changing the individual's attitudes, feelings, and thoughts is obvious. Educating the client about stuttering and the therapeutic process are essential for increasing the client's sense of self-control and therefore his fluency. This focus on both cognitive and motor characteristics of an individual who stutters relates closely to Fechner's work, in which mind and body are argued to be inseparable.

In Van Riper's approach to therapy, the client who stutters identifies and modifies the various thoughts and feelings he has regarding stuttering (e.g., I stutter because I am slow or I stutter because I have a nervous condition). The client achieves this by completing cognitive scales, which assess a variety of feelings and thoughts, throughout therapy. In parallel to completing such scales, the client participates in traditional motor skill drills-and-practiced exercises (e.g., appropriate air flow and smooth initiation of speech).

Stuttering therapy, according to Van Riper, facilitates individual changes in both cognition (feelings and thoughts about stuttering) and motor speech behaviors that decrease fluency. Both aspects of stuttering therapy, cognitive and motor speech, are necessary for developing natural, fluent speech. The relationship between cognition and motor speech control (fluency) is cyclical because as the client identifies/reduces his fears and negative thoughts he masters motor speech skills, which increase fluency.

Table 1: Components of therapeutic program for stuttering (Van Riper, 1973)

Therapy Phase	Clinician's Goal	Client's Goal	Therapy Tasks/Skills
Identification	-Analyze and classify characteristics of stuttering and client's reaction, attitudes -Identify ones own confidence in therapy techniques -Recognize ones own belief in change. -Identify for the client strengths and weaknesses	-Know his stuttering patterns within phrasal units -Recognition of stutter fear, speech fear, and generalized fear and anxiety	-Assessment -Diagnostic therapy
Desensitization	-Enables client to stutter without fear -Educates client about stuttering and dispels myths -Counsels client about psychological reactions	Becomes desensitized to stuttering and to fluency pattern	-Feedback via videotaping performance -Pseudostuttering -Drill-and-practice exercises to desensitize specific sounds and words
Modification	-Teaches strategies to control fluency -Continues to educate client about stuttering and myths -Continues to counsel client about psychological reactions	Demonstrates proficiency in controlling stutter events, coping with fear	-Cancellation -Pull-outs -Easy onsets -Motor planning techniques -Extinguish patterns that mimic or give the impression of fluency
Stabilization	-Steps into consultant role -Prepare client for termination of therapy, Relapse, old fears	-Sets his/her own therapy tasks, Identify success	-Develop proprioceptive monitoring of normal speech -Automatize motor planning strategies -Target any remaining stuttering behaviors

The Effectiveness of Cognitive Scales in Stuttering Therapy

This section presents data from therapists' and clients' ratings from cognitive scales at the beginning and termination phases of therapy.

The data that are presented below were gathered during the beginning and the termination of an intensive 75-hour, 10-week stuttering program with a group of 8 clients. Therapy took place three times a week and each session lasted for 2.5 hours. Along with the cognitive scales, clients' awareness regarding their speech was intended to increase by having them rate a variety of their motor/speech features. Clients rated their attitudes and/or various aspects of their speech during the sessions by completing various scales, which were individualized; based on what areas the client needed to improve. The same scales were rated by graduate clinicians in order to examine any differences between their ratings and the clients'.

The first client (Client 1) is a male with mild to moderate stuttering and the second client (Client 2) is a male with moderate to severe stuttering. Figures 1-4 show the differences in the ratings between each client and the clinician at the beginning and the end of therapy. In all of the figures, the features from the cognitive scale was used and the score range is shown as 0-7, with 7 being the best score for the specific category (e.g., no avoidance).

Overall, the Figures show that the scores improved, indicating progress in therapy. Figure 2 shows that the client's and clinician's ratings are the same in 5/8 measured attitudes and considerably dissimilar (a difference of 2 points) in two of the measured attitudes.

Figure 1: Ratings from Client 1 and Clinician at the Beginning of Therapy

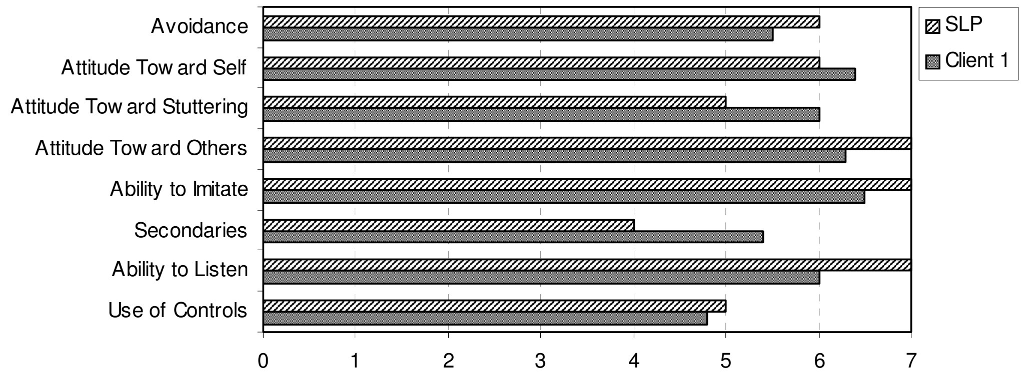


Figure 2: Ratings from Client 1 and Clinician at the End of Therapy

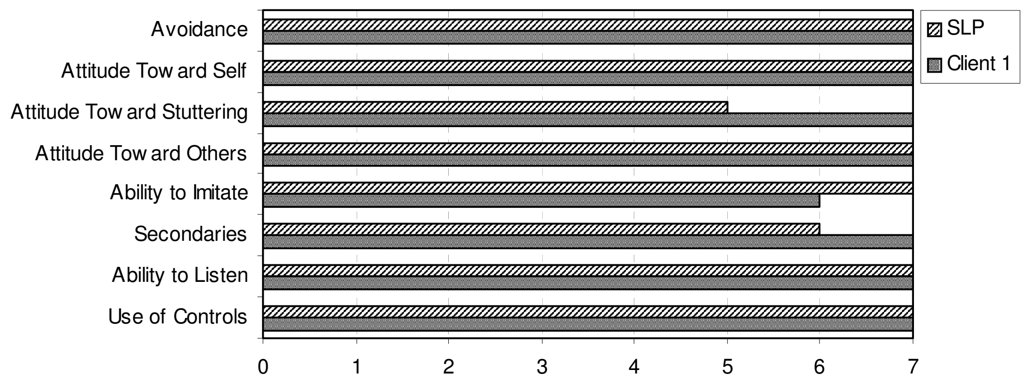


Table 2: Differences in Ratings by Clients and Clinicians at the Beginning and End of Therapy

Client 1	Avoidance	Attitude Toward Self	Attitude Toward Stuttering	Attitude Toward Others	Ability To Imitate	Second aries	Ability To Listen	Use Of Controls
Beg.	0.5	0.4	1	0.7	0.5	1.4	1	0.2
End	0	0	2	0	1	1	0	0
Client 2								
Beg.	1.2	2.8	2.8	0.8	0.8	0	0.3	0.4
End	2	2	0	1	0	1	1	3

However, the data is different in the case of Client 2, who has a moderate-severe stuttering. Overall, Figure 4 shows that the client did not score as highly as Client 1, suggesting that Client 2 had difficulty improving his attitudes and his fluency. Both Figures 3 and 4 demonstrate a noticeable difference between the clinician and the client.

Consequently, the data show that it is more likely for a client with mild-moderate stuttering to reach a high level of cognitive self-awareness. This is likely to help him use the fluency enhancement techniques in real life situations and control his fluency, based on the connection between cognition and body. On the other hand, a client with more severe stuttering will develop cognitive awareness at a lower degree. Additional research is needed to examine whether the difference in the degree of cognitive awareness is related to a similar degree of fluency control/fluent speech.

Figure 3: Ratings from Client 2 and Clinician at the Beginning of Therapy

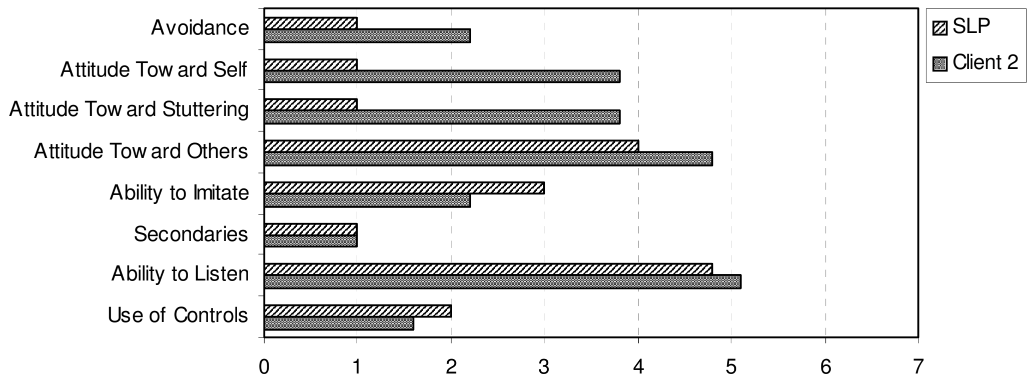
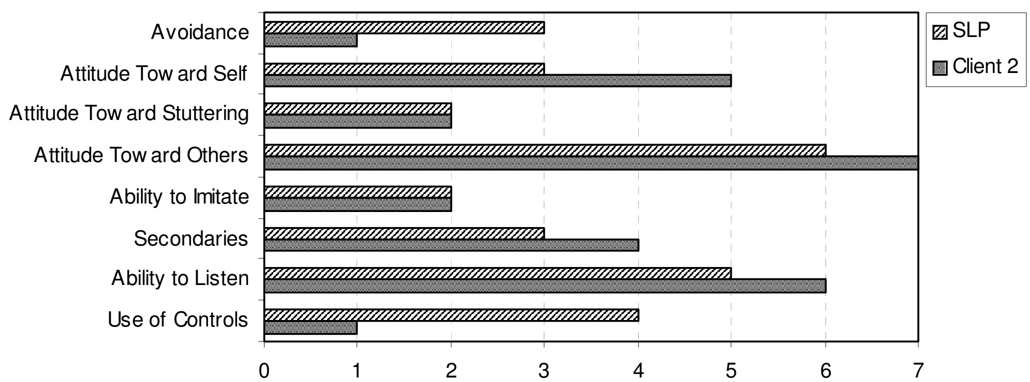


Figure 4: Ratings from Client 2 and Clinician at the End of Therapy



REFERENCES

Agnello, J. G. (1987). A Comprehensive program for facilitating fluency in stutterers. In Peters & Hulstijn (eds.), *Speech Motor Dynamics in Stuttering.* Springer-Verlag: New York.

Van Riper, C. (1971). *The nature of Stuttering*. Prentice-Hall: Englewood Cliffs, N.J.

Van Riper, C. (1973). *The Treatment of Stuttering*. Prentice-Hall: Englewood Cliffs, N.J.

Wingate, M. E. (2002). *Foundations of Stuttering*. Academic Press: San Diego, CA.

Wozniak, R. H. (1995). Mind and Body: Rene Déscartes to William James [online]. Available: http://serendip.brynmawr.edu/Mind/ (07/22/05).

A PSYCHOPHYSICIST EXAMINES AT THE NOTORIOUS STROOP EFFECT

Daniel Algom
Department of Psychology, Tel-Aviv University
algomd@freud.tau.ac.il

Abstract

The Stroop effect, psychology's classic measure of selectively and cognitive science's single most popular phenomenon, is shown to depend on a set of contextual factors well known in psychophysics. The effect is malleable, to the point of elimination, by systematic manipulations of these contexts. Consequently, the robustness of the effect is more apparent than real, result of systematic biases imputed into the experimental design through context.

The Stroop effect (Stroop, 1935) is psychology's classic measure to test the selectivity of attention to a particular aspect of a complex stimulus. When people name the color in which color words are printed their performance is impaired by the meaning of the carrier words. Word reading, in contrast, is not similarly hindered by irrelevant print color. The tenacity of the asymmetry in interference rendered the Stroop effect the demonstration of choice by instructors of psychology to illustrate the failure of selective attention. However, the import of the Stroop phenomenon exceeds classroom popularity. The effect has accrued theoretical interest because, presumably, it is the inescapable outcome of pitting the automatic process of word reading against the less automatic, or controlled, process of color naming. Given the imbalance in automatic activation and attendant speed of processing, the mandatory failure of selective attention to print color ensures. Consequently, the Stroop effect has been conferred the title, "gold standard" of attentional processing (MacLeod, 199), although a consensual theoretical interpretation proved elusive. It is for these reasons that more than 60 years after its discovery, the Stroop continues to fascinate researchers, sustaining an impressive amount of research. Indeed, the Stroop effect is the single most popular phenomenon studied in current cognitive science.

In the present analysis, I challenge the alleged robustness of the Stroop phenomenon as well as its traditional explanation in terms of automatic dominance of word over color. I show that the effect is systematically malleable by a set of (heretofore neglected) contextual factors. By judiciously manipulating these factors, one is able, at will, to fabricate a Stroop effect, to eliminate the effect, or reverse the effect (and create a situation in which word is impaired by color more than vice versa). Therefore, the effect is neither robust nor, indeed, inevitable. The analysis, in turn, casts doubt on traditional theories positing the automatic activation of word meaning as the root cause of the effect.

Relative Dimensional Discriminability

Chief among the recently identified contexts is the relative salience of the colors and words presented for view. Discriminability refers to the psychological differences separating stimulus values along the color dimension, on the one hand, and along the word dimension, on the other hand. Discriminability is matched when the print colors used are as easily (or as laboriously) distinguished from one another as are the color words used. Throughout large portions of the Stroop literature, relative dimensional discriminability has been neither appreciated nor measured. When it was, Algom, Melara, and their associates (e.g., Algom, Dekel, & Pansky, 1996; Algom, Chajut, & Lev, 004; Melara & Algom, 003; Melara & Mounts, 1993; Sabri, Melara, & Algom, 001) reproduced the classic Stroop asymmetry when the words were more salient than the colors. This word-color mismatch is characteristic of most published research with Stroop stimuli. Presenting, without precautions, the most widely available fonts and print colors is bound to yield just this mismatch. As a result, words might have intruded on colors not because they are automatically activated, but simply because they are more salient to perception. In general, a more salient dimension intrudes on performance with a less salient dimension more than vice versa.

When care was taken to make the colors as salient as the words, the asymmetry in interference (often, the entire Stroop effect) disappeared. Strikingly, when we purposely made the colors more discriminable than the words, then a reverse Stroop effect emerged with color intruded on word reading.

One should realize that modification of dimensional discriminability is not the trivial manipulation by which one makes the words illegible or the colors imperceptible. On the contrary, color naming and word reading are accomplished with nearly perfect accuracy in all experiments. Changing discriminability does not alter psychophysical performance with the pertinent stimuli. All are identified, recognized, or discriminated well beyond the constrains posed by imperfect resolution or the difference threshold. Relative dimensional discriminability concerns attention, not perception. Modifications of discriminability make one of the dimensions more salient, and hence attention catching, than the other.

Another subtle point should also be appreciated. Discriminability is manipulated willy-nilly in every experiment. When specifying the stimuli (e.g., selecting the fonts and the colors used), the experimenter unwittingly determines relative dimensional discriminability. The main point to note is that the experimenter, by creating the stimuli, indirectly governs attentional processing and influences the outcome.

Dimensional Correlation

Dimensional correlation specifies the rule by which the colors and the words are conjoined to form the Stroop stimuli in the experiment; hence, dimensional correlation is coextensive with the experimental design. Stroop (1935) presented only incongruent or conflicting color-word stimuli in his study (none of the words appeared in its matching color). Given the 5 color words and 5 print colors used, this arrangement amounts to a negative correlation of –0. 5 between word and color. Modern studies include congruent (the word naming its print color) as well as incongruent stimuli in the design. However, the accepted practice is to present the two types in equal numbers. This practice is conductive to a positive correlation between color and word in the experiment.
Here is the makeup of the vast majority of published Stroop experiments. Four color words and their corresponding print colors are presented. Notice that in this 4 x 4 matrix there are 4 congruent stimuli and 1 incongruent stimuli. To satisfy the stipulation of equal experimental frequency, each of the four color words appears <u>nine</u> times in its matching color, but <u>three</u> times in each of the three mismatching colors. This arrangement creates two equal subsets of 36 congruent and 36 incongruent stimuli, but it does so at the cost of imputing a substantial positive correlation between color and word over the experimental trials. The conditional probability of a print color (word) given a word (print color) is not the same for all words (print colors).

Dimensional correlation is fatal for selective attention. If there is a correlation, then the nominally irrelevant word actually carries information about the relevant color. The word becomes predictive of the target color. Noticing the word first on a particular trial provides the participant with a better than chance probability of guessing the print color. By establishing a(ny) correlation between color and word, the experimenter virtually dictates the diversion of attention toward the irrelevant dimension. Because they attend to the task-irrelevant dimension (to maximize performance), the participants open themselves up to Stroop interference (or facilitation).

Dishon-Berkovits and Algom (000) have purposely manipulated dimensional correlation and found that, when values along the target dimension and the to-be-ignored dimensions were correlated over the experimental trials, large effects of Stroop emerged. However, when truly random allocation of values created zero correlation, the Stroop effect largely vanished. The Stroop effect seems to depend on the (often hidden) presence of a predictive relationship between the nominally irrelevant dimension and the target dimension.

Response Production

This context refers to the mode participants use to render their experimental responses. In most Stroop experiments, the response mode was either vocal (orally naming the print color) or manual (pressing the appropriate key), although many other forms have also been used (e.g., card sorting, tapping, typing). There is an appreciable influence of response mode on the Stroop effect. For instance, vocal responding usually yields a larger effect than manual responding. Intuitively, response mode comes to affect the final stages of processing, those close to response emission, subsequent to stages of stimulus identification and decision making. Our research (Melara & Algom, 003; Sabri et al., 001) shows otherwise. We demonstrate that response mode affects decision evidence accumulation and, even earlier, stimulus identification and salience! It turns out that the effects of response mode are actually parallel to those of the psychophysical context of dimensional discriminability. Response mode acts to make one stimulus dimension more salient and accessible than the other, thereby commanding priority in the allocation of attention. For one example, the vocal response mode enhances the salience of the word dimension at the expense of the color dimension.

Set Size

Set size refers to the number of words and colors presented in the experiment. An experimenter can increase set size by including more values of color, more values of word, or both. As the experimenter adds dimensional values to the stimulus set, the probability of occurrence of any one value decreases, making its actual appearance on a trial more surprising. Increasing stimulus uncertainty of the task-irrelevant words is especially noteworthy. The larger the pool of words, the more informative is the word presented on any particular trial. Because the word is informative, it grabs attention at the cost of exclusive focusing on the relevant color. Applying principles of information theory explains the common observation that larger Stroop effects are obtained with larger sets of stimuli.

Overall Influence

Jointly, these four contexts shape performance in the Stroop task to a great degree. For instance, Melara and Algom (003) calculated for each of 34 experiments culled from the literature the correlation between (a) the difference in baseline performance with the colors alone and the words alone, i.e., relative dimensional salience, and (b) the Stroop effect obtained in that experiment. The Pearson correlation amounted to .78 between these two variables. In a similar vein, the authors also calculated (in the same studies) the correlation between (a) the correlation between color and word built into the experimental design, and the (b) the Stroop effect obtained. This correlation amounted to .69. Together, these two factors of context alone produced a multiple correlation of .87 with the Stroop effect. Despite the disparate methods, stimuli, and populations included in that sample of studies from the literature, discriminability and correlation accounted for more than 75% of the variance in the observed Stroop effects.

Why did Stroop and many subsequent researchers obtain the Stroop effect? On the present view, they introduced contexts that heavily favored word over color in the pertinent psychological experience. In Stroop's (1935) study, perceptual salience was biased in favor of word. Word naming (with words appearing in uniform black) was much faster than color naming (of irregular shapes). The production context, oral naming, acted further to enhance the salience of word at the expense of color. Color and word were correlated over the trials such that the latter was predictive of the former. And, the distractor word was moderately surprising on each trial. Collectively therefore, the experienced changes in word were enormous compared with those in color. The Stroop effect ensued. Theoretically, however, these forms of contexts are not fixed or inevitable. Removing the imbalance and attendant biases is possible (in fact, easy to accomplish) and results in a different outcome (often, in the elimination of the Stroop effect).

The basic tenet of the present contextual approach is this. The more salient and informative a distractor is about the target -- expressed, respectively, by discriminability and response production, and by correlation and set size -- the more likely it is that selective attention to the target will fail. According to a recent theory (Melara & Algom, 003), the four contexts alter two structural representations -- dimensional imbalance and dimensional uncertainty – and, in so doing, predictably affect the magnitude, the direction, and the very presence of the Stroop effect. Figure 1 gives a schematic representation of tectonic theory.

Tectonic theory, guided by insights from information theory, predicts naturally the full variety of Stroop and reverse Stroop effects. The theory stands alone in elucidating the pliability of the effect, denying a range of immutable properties attributed to the effect by traditional explanations.

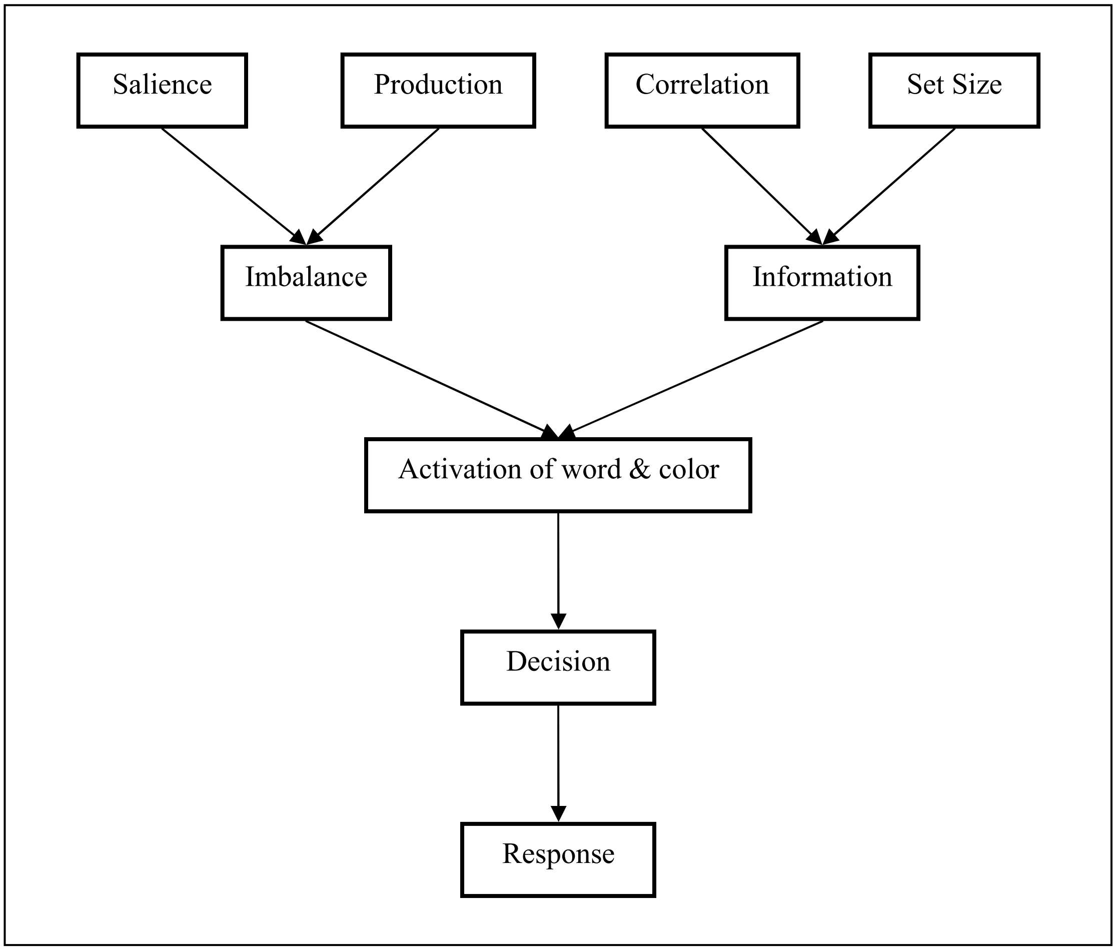


Figure 1: Tectonic theory of Stroop effects (Melara & Algom, 003). Four contexts under control of the experimenter or nature selectively influence two theorized structures, dimensional imbalance and dimensional information. The structures, in turn, govern the parameters of the activation functions for word and color. The activations are entered into a decision mechanism configured as evidence ratio for the target (color) response. A decision is made and response is emitted when the ratio reaches a preset threshold.

References

Algom, D., Chajut, E., & Lev, S. (004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General, 133,* 3 3-338.

Algom, D., Dekel, A., & Pansky, A. (1996). The perception of number from the separability of the stimulus: The Stroop effect revisited. *Memory & Cognition, 24,* 557-57 .

Dishon-Berkovits, M., & Algom, D. (000). The Stroop effect: It isnot the robust phenomenon that you have thought it to be. *Memory & Cognition, 28,* 1437-1449.

Macleod, C. M. (199). The Stroop task: The “gold standard” of attentional measures. *Journal of Experimental Psychology: General, 121,* 1 -14.

Melara, R. D., & Algom, D. (003). Driven by information: A tectonictheory of Stroop effects. *Psychological Review, 110*, 4 -471.

Melara, R. D., Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21,* 6 7-645.

Sabri, M., Melara, R. D., & Agom, D. (001). A confluence of contexts: Asymmetric versus global failure of selective attention to Stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 515-537.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions*. Journal of Experimental Psychology, 18,* 643-66 .

INDEPENDENT PARALLEL CHANNELS PREDICT THE STROOP EFFECT

Daniel Algom[1], Ami Eidels[2], James T. Townsend[2], and Helena Kadlec[3]
[1]*Tel Aviv University, algomd@freud.tau.ac.il*
[2]*Indiana University, aeidels@indiana.edu, jtownsen@indiana.edu*
[3]*University of Victoria, hkadlec@uvic.ca*

Abstract

We present a new model that explains the Stroop effect. In our model, each attribute of the colored word stimuli in a Stroop task (each word and each font color) is processed on a separate channel, but our model differs from all other models of the Stroop effect because it proposes that these channels are parallel and completely independent. Our simulated results with response times and proportions of correct trials show that our model easily predicts the standard behavioral Stroop results – colors are named faster and more accurately when the word is congruent with its font color than when the word is incongruent with its color. Furthermore, the model also accounts for contextual influences that have been observed with the Stroop effect, such as relative dimensional discriminability, dimensional covariation, set size, mode of responding and task difficulty.

In a Stroop task, words of different colors (e.g., the letters "RED" printed in a red or green colored font) are presented one at a time and the observer is asked to name the font color of the word as quickly and accurately as possible. Words that match their font color (e.g., the word "RED" in a red font) are called congruent and words that do not match their color (e.g., "RED" in a green font) are called incongruent. (Note that we will designate the word in capital letters and the font color will be lowercase and underlined.) The "standard" Stroop result is that the color-word interferes with (increases response times and lowers accuracy of) naming of an incongruent font color.

In looking for a compelling explanation of the Stroop effect, Algom and colleagues have recently examined the effect under a variety of experimental conditions and found that several contextual factors impact the magnitude of the behavioral Stroop effect (Melara & Algom, 2003). Specifically, when the relative discriminability of the words is better than of the colors (the typical experimental situation), the "standard" Stroop effect is found; however, when the relative discriminability is reversed by using more discriminable colors, the reverse Stroop effect appears (e.g., Algom, Chajut, & Lev, 2004). Another contextual factor that impacts the Stroop effect is the existence of dimensional covariation; in the typical experimental situation, congruent stimuli are presented more frequently than incongruent stimuli (to balance the number of word and color combinations). Again, when this factor is experimentally manipulated it affects the magnitude of the Stroop effect (e.g., Dishon-Berkovits & Algom, 2000). Furthermore, the magnitude of the Stroop effect is also influenced by the set size (a larger stimulus set leads to a larger Stroop effect; e.g., Kanne, Balota, Spieler, & Faust, 1998), the response mode (oral responding leads to an stronger Stroop effect than manual responding), and likely also task difficulty (easier tasks seem to produce a smaller Stroop effect). In light of these results, however, it is interesting to note that the Stroop effect is observed even when the words and colors used in the experiment are equally salient and randomly mixed and presented.

Explanations of the Stroop effect have centered on the *interactions* of the information processing of colors and words. In other words, most theoretical explanations have focused on how the word facilitates (in congruent cases) and interferes with (in incongruent cases) naming of the color. The model presented here, on the other hand, is unique in that makes the correct predictions without postulating any interactive effects. As we show below, our model with parallel and completely independent channels correctly predicts shorter response times and higher accuracy rates with congruent stimuli, and also accommodates the contextual influences described above.

The Separate Channels Model with Four Independent Parallel Accumulator Channels

We derive our model for the four-stimulus design with two color-words (RED, GREEN) combined factorially with two font colors (red, green). Let us denote each stimulus – a color-word printed in a particular font color – by the pair (*W,c*), where *W* indicates the stimulus word (*R* for "RED" and *G* for "GREEN") and *c* indicates the font color *(r* for red and *g* for green). E.g., (*R,g*) denotes the word "RED" printed in a green font. Also, recall that the participant's correct response is the name of the font color.

Processing Assumptions

1. All of the words and font colors included in the stimulus set are activated on each trial. For the 2(word) x 2(color) situation here, all four channels – RED, GREEN, red, and green – are activated on each trial of the experiment. In a different experiment, e.g., given a Stroop matrix of 6 colors and 10 words, 16 channels are activated on each trial.
2. The architecture is that of a strictly parallel, independent race. The channels do not converge downstream, nor does there exist a common decision mechanism fed by the various channels. Each channel processes its own input and races to produce a response determined entirely by the information it carries.
3. Processing is stochastic, accomplished with imperfect accuracy. In our four-channel model, we assume that each channel accumulates information and the correct response is produced when the channel entailing the presented font color or the channel entailing its matching color word (whether or not that word is presented on that particular trial) wins the race and determines the response. An error occurs when any of the other channels wins the race. Errors are relatively infrequent due to consistent discouraging feedback from the experimenter (or nature).
4. Attributes exposed for view on a particular trial are processed more efficiently than the non-exposed attributes. For example, when stimulus (*G,r*) is presented, the *G* and *r* counters will accumulate counts at a faster rate than the *R* and *g* counters.
5. (Optional). The font-color channels can be faster than the word channels. This assumption reflects compliance with the experimental instructions because the participant is instructed to attend to the font color. When the Stroop experiment includes the complementary condition of word reading, then, the words comprise the relevant dimension and a reverse Stroop effect (by which the irrelevant font colors interfere with word reading) is often observed. This assumption is optional because the first four suffice to produce the behavioral Stroop effects.

Formal Model

The formal description of the four independent-parallel-channel model draws heavily on the work by Townsend and Ashby (1984, pp. 272-279). It is a basic counter model with the accumulation rates modeled by a Poisson distribution (as per assumption 3 above). This means that the time interval for the arrival of the next unit of information (count) on a given channel *m* has an exponential distribution with a rate parameter λ_m, viz., $f(t)=\lambda_m \exp(-\lambda_m t)$. For our four-channel model, we have one channel for each stimulus attribute (i.e., R, G, r, g) and because there are four possible stimuli in the experiment, we must also distinguish the processing rates (for each channel) for each of the four stimuli. We therefore denote $\lambda_{(R,r),R}$ as the rate parameter of the R channel when the stimulus (R,r) was presented. Our model is shown schematically in Fig. 1. The length of the arrow in each stimulus panel represents the relative magnitude of the corresponding rate parameter. For example, for stimulus *(R,g)*, the longest arrow is for channel g, hence $\lambda_{(R,g),g}$ is the largest rate parameter for this stimulus indicating that this channel is most likely to finish processing first on most trials where *(R,g)* was presented. Upon the display of a stimulus, processing occurs on all four channels (as per processing assumption 1 above) until a criterion, K_m, is reached by one of the processing channels m, at which point response m is emitted.

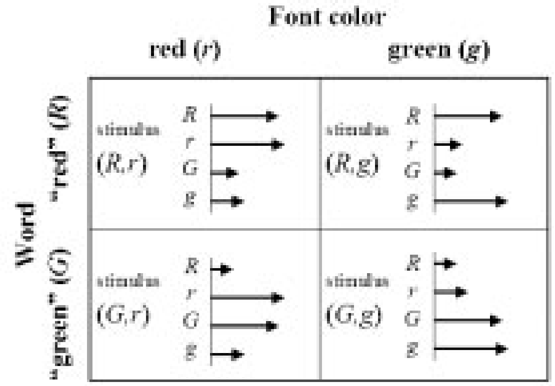


Fig. 1. Separate channels model with 4 parallel and independent channels. See text for explanation.

The full 4-channel model has a total of 20 parameters: 16 rate parameters (one for each of the 4 counters and each of the 4 stimuli in the set) and four counter criteria, one for each of the four counters – K_R, K_r, K_G, K_g. (We assume that these don't differ across the four stimuli.) We can, however, make several sensible simplifying assumptions, namely, (1) we set all four counter criteria to the same value: $K_R = K_r = K_G = K_g = K$; and (2) we set the rate parameters as follows:

$\lambda_{(R,r),r} = \lambda_{(R,g),g} = \lambda_{(G,r),r} = \lambda_{(G,g),g} = \lambda_{CC}$. (where CC is the **c**onsistent **c**olor channel);
$\lambda_{(R,r),R} = \lambda_{(R,g),R} = \lambda_{(G,r),G} = \lambda_{(G,g),G} = \lambda_{CW}$ (where CW is the **c**onsistent **w**ord channel);
$\lambda_{(R,r),g} = \lambda_{(R,g),r} = \lambda_{(G,r),g} = \lambda_{(G,g),r} = \lambda_{IC}$ (where IC is the **i**nconsistent **c**olor channel);
$\lambda_{(R,r),G} = \lambda_{(R,g),G} = \lambda_{(G,r),R} = \lambda_{(G,g),R} = \lambda_{CW}$ (where IW is the **i**nconsistent **w**ord channel).

The five processing assumptions lead to the ordering on the (simplified) rate parameters as

$$\lambda_{CC} > \lambda_{CW} > \lambda_{IC} > \lambda_{IW}. \tag{1}$$

Predictions and Results for Response Times

The mean response time to respond "red" for each of the four stimuli (W,c), denoted by $T_{\text{"red"}|(W,c)}$, are given by:

$$T_{\text{"red"}|(R,r)} = \frac{1}{P[\text{"red"}\,|\,(R,r)]}\sum_{j=0}^{2K-2}\binom{K+j-1}{j}\left(\frac{\lambda_{IW}+\lambda_{IC}}{\sum\lambda}\right)^{j}\left(\frac{\lambda_{CW}+\lambda_{CC}}{\sum\lambda}\right)^{K}\left(\frac{K+j}{\sum\lambda}\right) \tag{2a}$$

$$T_{\text{"red"}|(R,g)} = \frac{1}{P[\text{"red"}\,|\,(R,g)]}\sum_{j=0}^{2K-2}\binom{K+j-1}{j}\left(\frac{\lambda_{IW}+\lambda_{CC}}{\sum\lambda}\right)^{j}\left(\frac{\lambda_{CW}+\lambda_{IC}}{\sum\lambda}\right)^{K}\left(\frac{K+j}{\sum\lambda}\right) \tag{2b}$$

$$T_{\text{"red"}|(G,r)} = \frac{1}{P[\text{"red"}\,|\,(G,r)]}\sum_{j=0}^{2K-2}\binom{K+j-1}{j}\left(\frac{\lambda_{CW}+\lambda_{IC}}{\sum\lambda}\right)^{j}\left(\frac{\lambda_{IW}+\lambda_{CC}}{\sum\lambda}\right)^{K}\left(\frac{K+j}{\sum\lambda}\right) \tag{2c}$$

$$T_{"red"|(G,g)} = \frac{1}{P["red"|(G,g)]} \sum_{j=0}^{2K-2} \binom{K+j-1}{j} \left(\frac{\lambda_{CW}+\lambda_{CC}}{\sum \lambda} \right)^{j} \left(\frac{\lambda_{IW}+\lambda_{IC}}{\sum \lambda} \right)^{K} \left(\frac{K+j}{\sum \lambda} \right) \quad (2d)$$

The Stroop effect is given by comparing Eqs. (2a) and (2c) to see if $T_{\text{"red"}|(R,r)} < T_{\text{"red"}|(G,r)}$. It turns out that:

> ***Proposition 1:*** In a 4-parallel independent channel model, $T_{\text{"red"}|(R,r)} < T_{\text{"red"}|(G,r)}$ whenever $\lambda_{CW} > \lambda_{IW}$ for any value of $K>1$. [For K=1, $T_{\text{"red"}|(R,r)} = T_{\text{"red"}|(G,r)}$.]

We have an analytic proof for K=1, 2 and 3 (available on request), and Fig. 2 shows the simulated results for a large number of λ parameter values and selected values of K. In all cases, the response time for the congruent stimulus was faster then for the incongruent stimulus (all points fall below the diagonal line where the RT for congruent is equal to the RT for the incongruent stimulus).

Predictions and Results for Accuracy

The probabilities of responding "red" when each of the four stimuli is presented are given by:

$$P["red"|(R,r)] = \sum_{j=0}^{K-1} \binom{K+j-1}{j} \left(\frac{\lambda_{IW}+\lambda_{IC}}{\sum \lambda} \right)^{j} \left(\frac{\lambda_{CW}+\lambda_{CC}}{\sum \lambda} \right)^{K} \quad (3a)$$

$$P["red"|(R,g)] = \sum_{j=0}^{K-1} \binom{K+j-1}{j} \left(\frac{\lambda_{IW}+\lambda_{CC}}{\sum \lambda} \right)^{j} \left(\frac{\lambda_{CW}+\lambda_{IC}}{\sum \lambda} \right)^{K} \quad (3b)$$

$$P["red"|(G,r)] = \sum_{j=0}^{K-1} \binom{K+j-1}{j} \left(\frac{\lambda_{CW}+\lambda_{IC}}{\sum \lambda} \right)^{j} \left(\frac{\lambda_{IW}+\lambda_{CC}}{\sum \lambda} \right)^{K} \quad (3c)$$

$$P["red"|(G,g)] = \sum_{j=0}^{K-1} \binom{K+j-1}{j} \left(\frac{\lambda_{CW}+\lambda_{CC}}{\sum \lambda} \right)^{j} \left(\frac{\lambda_{IW}+\lambda_{IC}}{\sum \lambda} \right)^{K} \quad (3d)$$

Stroop effect for performance accuracy is given by comparing Eqs. (3a) and (3c), to determine if *P["red"|(R,r)] > P["red"|(G,r)]*. And as we saw with the response times, indeed we have:

> ***Proposition 2:*** In a 4 parallel-independent-channel model, *P["red"|(R,r)] > P["red"|(G,r)]* whenever $\lambda_{CW} > \lambda_{IW}$ for any value of K.

Analytic proof for K=1,2 and 3 is available on request, and simulated results are shown in Fig. 3. For all cases, we see that accuracy is better for the congruent than for the incongruent stimuli.

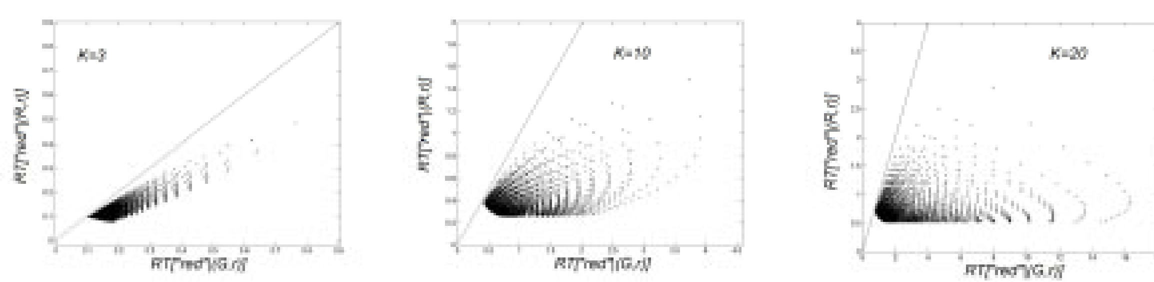


Fig. 2. Simulation results for response times for K=3, 10 and 20. In each case, the congruent stimulus leads to a faster response time (below the line) than the incongruent stimulus.

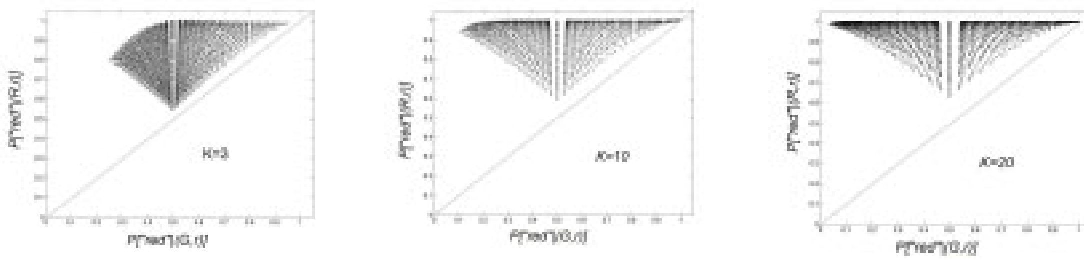


Fig. 3. Simulation results for accuracy for K=3, 10 and 20 and a wide range of λs ordered as given in Eq. (1). In each simulated case, the congruent stimulus leads to a more accurate rate of responding than the incongruent stimulus.

Contextual Influences

We mentioned in the introduction that the magnitude of the Stroop effect is affected by several contextual factors. Here, we briefly outline how the proposed model accounts for these influences.

(1) The effect of different relative discriminability of the words and colors is captured in the model by increasing the rate parameters for the R channels for RED stimuli, $\lambda_{(R,r),R}$ and $\lambda_{(R,g),R}$, and similarly increasing G channels for stimuli (G,g) and (G,r), as depicted in Fig. 4. In this situation, the rate parameters have the ordering $\lambda_{CW} > \lambda_{CC} > \lambda_{IW} > \lambda_{IC}$, and as before the congruent stimuli are both faster and more accurate than the incongruent stimuli.

(2) To explain the enhanced Stroop effect with correlated dimensions, we first note that the "standard" Stroop effect described above using the difference between Eqs (3a) and (3c) – viz. $\Delta P = P[\text{"red"}|(R,r)] - P[\text{"red"}|(G,r)] > 0$ – holds when the congruent and incongruent stimuli are presented with the same frequency; i.e., when $P[(R,r)]=P[(G,r)]$. But $P[\text{"red"}|(W,c)]$ is the probability of responding "red" on **one** trial of stimulus (W,c), and the Stroop effect is obtained over a large number of trials. Hence we should really write the Stroop effect as the difference $\Delta P = P[\text{"red"} \cap (R,r)] - P[\text{"red"} \cap (G,r)] > 0$. Using $P(A \cap B) = P(A|B)\ P(B)$ gives $\Delta P = P[\text{"red"}|(R,r)]\ P[(R,r)] - P[\text{"red"}|(G,r)]\ P[(G,r)]$. Note that with $P[(R,r)]=P[(G,r)]$, the two ways of writing ΔP, using conditional probabilities or joint probabilities, are identical. With correlated dimensions, we have $P[(R,r)]>P[(G,r)]$, and after several algebraic step we obtain the result that $\Delta P_{correlated} - \Delta P_{standard} > 0$, hence showing an enhanced Stroop effect for the correlated condition.

(3) As set size of the stimulus ensemble increases, the relative frequency of congruent stimuli decreases. For example, congruent combinations form a quarter of the stimuli in a 4 x 4 matrix, but they form a mere 12.5% of the stimuli in a 8 x 8 matrix. Because congruent and

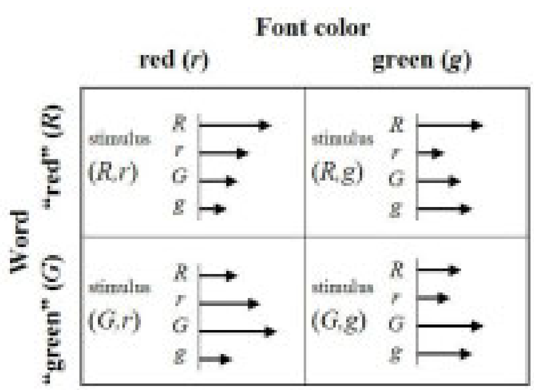


Fig. 4. Separate channels model for dimensional discriminability that favors WORD channels R and G over color channels r and g.

incongruent stimuli are nonetheless matched in experimental frequency, set size magnifies the asymmetry already noted with respect to the covariation context in (2) above.

(4) To account for an enhanced Stroop effect with an oral mode of responding, we note that oral responding increases the tendency of substituting the corresponding word channel for the color channel, thus enhancing the Stroop effect. And finally,

(5) A difficult Stroop task augments the frequency of the matching word channels, augmenting, in turn, the size of the observed Stroop effect.

Discussion

The conditions enabling the behavioral Stroop effect in the context of our new model are as follows. For the congruent stimulus (e.g., RED in red), both the font color red and the color word RED are presented for view and count for the correct response. For the corresponding incongruent stimulus (GREEN in red), the presented color red and the unexposed word RED count for the correct response. The RED channel thus carries perceptual information in the first case, and only non-perceptual information in the second case. According to processing assumption 4, however, perceived stimulus attributes are processed more efficiently than non-exposed attributes. Therefore, the RED channel is processed more efficiently with congruent than with incongruent stimuli. Hence we observe the Stroop effect – overall better performance with congruent stimuli over repeated trials.

Furthermore, according to the third processing assumption, processing is stochastic, and various channels apart from the presented font color can win the race (and thus completely determine the response) on the various trials. The most important channels in the Stroop effect are those that carry the matching words, perceptually (when presented for view) or from memory (when another word is presented). And while an erroneous response ensues when any of the other channels wins the race, this would be a rare event given the pervasive negative feedback. Thus on our separate channels theory, the Stroop phenomenon is parasitic on stochastic processing; there would be none under deterministic processing.

Finally, as we stressed several times, the stimulus attributes are processed in separate and independent channels (processing assumption 2). As a result, the channel carrying the presented font color (i.e., the component selected for responding by the experimental instructions) is indifferent to the channel carrying the presented word (as well as to all other channels). What this means is that when the color channel determines the response, the response times and accuracy are the same whether the stimulus was congruent or incongruent. Thus a provocative implication of this model is that congruity itself does not carry psychological reality in the parallel race!

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *J of Experimental Psychology: General, 133,* 323-338.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory & Cognition, 28*, 1437-1449.

Kanne, S.M., Balota, D.A., Spieler, D.H., & Faust, M.E. (1998). Explorations of Cohen, Dunhar, & McClelland's (1990) connectionist model of Stroop performance. *Psych. Rev, 105*, 174-187.

Melara, R. D. & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychol. Review, 110*, 422-471.

Townsend, J.T., & Ashby, F.G. (1983). *The Stochastic Modeling of Elementary Psychological Processes*. NY: Cambridge U Press.

MOTION ACCELERATION AFFECTS THE LOCALIZATION OF THE VANISHING (BUT NOT OF THE STARTING) POINT

Alessia Bastianelli, Rossana Actis-Grosso and Natale Stucchi
Dipartimento di Psicologia, Università di Milano-Bicocca, Italy.
e-mail: alessia.bastianelli@.unimib.it.

Abstract

Observers make localization errors when asked to localize the perceived onset (starting point, SP) and offset (vanishing point, VP) of a moving target. Four experiments are presented, aimed to test the influence on SP and VP errors of the modulation of target's velocity in the first- (Experiment 1) and the last-part (Experiment 2) of a 12.5° path. Consistently with previous results, a forward VP displacement was obtained, whereas for SP a backward displacement proved significant only for horizontal motions. Results showed an effect of velocity modulation on VP but not on SP. Two controls experiments were performed, in which target's constant velocity corresponded to the instantaneous velocity it had at SP (Experiment 3) and at VP (Experiment 4). The trend of results was the same as in the main experiments, thus supporting the hypothesis that a low level mechanism is responsible for the forward displacement of VP.

There have been several reports of mislocations associated with the onset and the offset of moving objects.

- For the onset (starting point, SP), a localization error in the direction of motion (i.e., a forward mislocation) was firstly reported by Fröhlich (*Fröhlich effect*, 1923), whereas recent studies also revealed the reverse error, that is an error opposite to the direction of motion (i.e., a backward mislocation; Actis-Grosso, Stucchi and Vicario, 1996; Hubbard and Motes, 2002; Thornton, 2002). For the Fröhlich effect, some partially contradictory accounts have been proposed, essentially based on attentional mechanisms (Müsseler and Aschersleben, 1998; Kirschfeld and Kammer, 1999). For the backward mislocation, a hypothesis have been suggested by Actis-Grosso and Stucchi (2003), based on an early visual extrapolating mechanism (Nijhawan, 1994) that compensates the ~100 ms neural delay by transforming it into a spatial distance. This transformation could be exaggerated at the beginning of motion, since an object moving at a constant speed is seen as accelerating at the beginning (Runeson, 1974). According to this hypothesis, target's absolute velocity at the beginning of motion is a crucial factor (the mechanism is supposed to have a threshold above which it cuts the first 100 ms of motion), together with the presence of an anchor. If a static anchor is present in the visual scene, such as an occluding surface from which the target starts its motion, the expected error should always be in the direction of motion, independently on target's velocity. Other accounts are based on trial context (and in particular on predictability, Müsseler and Kerzel, 2004) and on an attentional bias relative to motor judgements (Kerzel and Gegenfurtner, 2004).

- For the offset localization (vanishing point, VP) an error in the direction of motion has been reported (Scholz, 1924; Hubbard and Bharucha, 1988): this error was referred to as "representational momentum" by Hubbard (e.g. 1990), and for it different interpretations were put forth, based on memory (e.g. Hubbard, 1995), attention (Kerzel, 2003a) and perception (Kerzel et al., 2001). In particular, Hubbard suggested that a high level cognitive mechanism anticipates the event course on the basis of the knowledge of the previous pattern of behaviour of the moving object. Verfaille and d'Ydewalle (1991) proposed that the event course anticipation occurs on the basis of the representation of a higher-level structure of the motion event as a whole rather than on the basis of the local, ongoing motion trajectory. Recently, both Kerzel (2003b) and Actis-Grosso and Stucchi (2003) have proposed explanations for VP displacement not directly referred to higher-level cognitive processes, albeit both explanations admit a possible intervention of expectations and memory on the localization error.

This work is a part of a research aimed to test the role of velocity modulation on both SP and VP mislocations, where in different experiments absolute velocity and modulation of velocity are studied at the beginning, in the central part and in the final part of a motion. Last year we presented data concerning the modulation of velocity in the central section of a motion trajectory (Bastianelli et al., 2004); here we will discuss data concerning the initial (Experiment 1) and final (Experiment 2) sections. In particular, if the model proposed by Actis-Grosso and Stucchi (2003) is correct, we should expect that the modulation of velocity, in the initial part of a motion has an effect on the SP error: a deceleration should reduce the error, an acceleration should increase it.

Experiments 1 and 2

Method

Participants. Ten naïve participants (seven females and three males, average age 23.6 years) volunteered for the experiment. All had normal or corrected-to-normal vision.

Apparatus and stimuli. The stimuli were presented on a PC Pentium 4- based computer equipment connected to a 21 in. monitor with a resolution of 1600 x 1200 pixels (where a pixel can be considered as a square of 0.1 mm) and a refresh rate of 100 Hz. A red dot (5 pixels in diameter) travelled on a black background in a straight line covering a path of 12.5 cm in ~1.8s. The two horizontal directions were right-to-left (RL) and left-to-right (LR); the two vertical directions were top-to-bottom (TB) and bottom-to-top (BT). For 1/5 of the stimuli, the velocity of the disc was kept constant at 6.37 cm/s, whereas for the remaining stimuli it was varied following one of four different modulations. The dot could have four different instantaneous velocities at SP (Experiment 1) and at VP (Experiment 2); in Experiment 1 it reached the velocity of 6.37 cm/s in the first 482 ms of motion, in Experiment 2 it went from 6.37 cm/s to the instantaneous final velocity in the last 482 ms of motion. In this way, we obtained two linearly accelerated and two linearly decelerated motions, as listed below (Experiment 1).

Instantaneous starting velocity	Corresponding acceleration value	Code
2.12 cm/s	8.82 cm/s^2	+2
4.25 cm/s	4.41 cm/s^2	+1
6.37 cm/s	0	0
8.50 cm/s	-4.41 cm/s^2	-1
10.62 cm/s	-8.82 cm/s^2	-2

These values corresponded to the final velocities in Experiment 2 and then the corresponding the acceleration values – and the relative codes - were inverted (e.g., if the instantaneous final velocity was 2.12 cm/s, the corresponding acceleration value was -8.82 and consequently it was coded as -2).

The actual SP could be in a random position varied between 5 and 7.5 cm from the centre of the screen. Stimuli were randomized between subjects.

Procedure. Participants sat at a comfortable distance from the screen (about 60 cm) in a dark room. A blue dot, having the same dimension of the target (5 pixels in diameter) signalled the point at which the stimulus was going to appear. The participant was not told that the position of the attention signal coincided with the SP of motion. The blue dot remained on the screen for 500 ms, followed by a blank interval randomly chosen between 1 and 2 s, and then the red dot appeared, and the motion started at once. At the end of motion, the participants had to localize by using the mouse where the moving disk had appeared. As soon as the mouse was moved, a white cross-hair (3 x 3 pixels) was displayed on the screen in a random position comprised between 3 and 4.5 cm from the actual SP (Experiment 1) or the actual VP (Experiment 2). The task of participants consisted of placing the cross exactly where he/she had seen the disk to appear (Experiment 1) or to disappear (Experiment 2) and pressing the mouse button. The response marked the start of the next trial. The completely within-subjects experimental design included two factors: Direction (4 levels) x Modulation (5 levels) for a total of 20 stimuli. Each stimulus had five replications, for a total of 100 trials. The duration of one experiment was approximately 16 minutes. The order of presentation of the two experiments was counterbalanced between participants.

Results

The difference between judged starting (vanishing) point and actual starting (vanishing) point is referred to as displacement. For horizontal motion, we analysed only the *x* displacement and for vertical motion only the *y* displacement. The mean constant error (CE) was obtained by subtracting the coordinate of the response from the actual coordinate, thus obtaining a negative value for backward displacement and a positive value for forward displacement.

SP. In figure 1 mean errors are reported for the four directions of motion. Consistently with previous results, a backward SP displacement was obtained. The average error in pixels computed over participants (-7.27) proved not significant. As it is clear from Figure 1, this is mainly due to vertical direction. Both horizontal directions proved significant [CE=-17.04, SE=5.29 $t(9)=7.28$, $p<.000$, for LR, and CE=-13.67, SE=5.98 $t(9)=2.75$, $p<.01$ for RL respectively] when the mouse coordinates were confronted with the actual coordinates.

A mirror reflection for the Ces of SP was needed to perform the ANOVA on CE. Because ANOVA is aimed at detecting systematic effects in the magnitude of CE, the sign of CE of SP and VP should be changed to obtain positive values for both backward displacement of SP and forward displacement of VP.

A two-way ANOVA on CE showed only an effect of Direction [$F(1,9)=3.9$ $p< .02$]. This effect is due to the difference between horizontal (CE=-15.35 SE=3.80) and vertical directions [CE=0.79, SE=5.22 $t(9)=-4.607$, $p<.001$]. In fact, for constant velocity profile, the error was statistically significant for both horizontal directions [(CE=-16.56, SE=4.18) $t(9)=2.43$, $p<.01$ for LR and CE=-13.23, SE=5.60 $t(9)=-3.12$, $p<.003$ for RL respectively]. This substantiates previous results. Furthermore, for LR direction, both positive accelerations proved significant [(CE = -14.62, SE=3.52) $t(9)=3.64$, $p<.001$ for +2 and CE=-10.67, SE=6.54 $t(9)=2.79$, $p<.01$ for +1 respectively], while for RL direction both negative accelerations proved significant [(CE=-14.4, SE=6.49) $t(9)=-2.54$, $p<.02$ for -2 and CE=-8.25, SE=7.15 $t(9)=3.29$, $p<.002$ for -1 respectively].

The absence of a significant error in the vertical direction, albeit consistent with our previous results, is not consistently reported in the literature and needs further investigation.

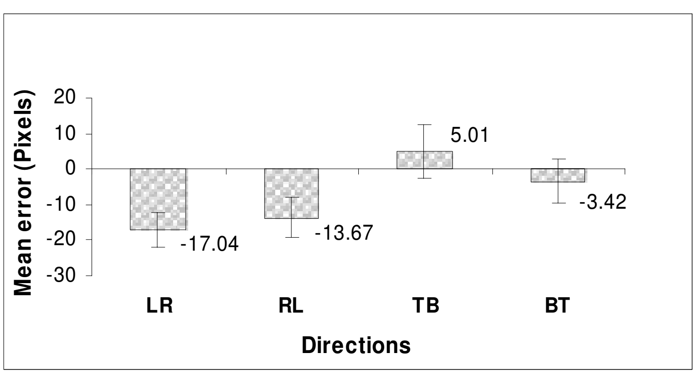


FIGURE 1. Mean errors for the four directions of motion for SP. Standard errors on top.

VP. In figure 2 mean errors for VP are reported for the five modulation of velocity. Consistently with previous results, a forward VP displacement was obtained.

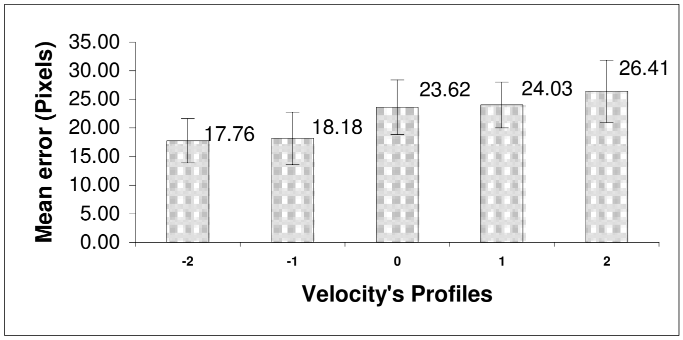


FIGURE 2. Mean errors for the five velocity modulations for VP. Standard errors on top.

The average error in pixels computed over participants was 23.23 (SD= 13.94) [t(9) = -2.86 p<.001]. VP error proved significant for all velocity's profiles and directions except than for the constant velocity profile and the two positive accelerated profiles in the direction RL.
A two-way ANOVA on CE showed an effect of Acceleration [F(1,9)=5.5 p<.002]: VP error increased as a function of Acceleration. This does not necessarily mean that velocity modulation has an effect on VP error, because the CE increased also as a function of final target's instantaneous velocity, and this could simply be due to the well known fact that target's absolute velocity influence the magnitude of the error (e.g. Hubbard, 1995). A control experiment is thus necessary to verify whether this result is due to the modulation of velocity or to the instantaneous final velocity of the target.

Experiments 3 and 4 (control experiments)

Method

Participants, apparatus, stimuli and procedure were the same as in previous experiments, with the following exception: target's constant velocity (4 levels) corresponded to the instantaneous velocity it had at SP in Experiment 1 and at VP in Experiment 2. Thus, four constant velocity profiles were used, identical for both Experiment 3 and 4, corresponding to 2.12, 4.25, 8.5 and 10.62 cm/s. Participant's task was to move the mouse cursor where he/she saw the target to appear (Experiment 3) or to disappear (Experiment 4).
The order of presentation of the two experiments was counterbalanced between participants.

Results

Results of experiments 3 and 4 substantially confirmed main experiments. A backward SP displacement (CE=-10.24 pixels) and a forward VP displacement [(CE=-28.1 pixels, SD= 32.36) t(9) = 8.12 p< .001] were found. As in the main experiments, total CE was not significant for SP, but it was significant for both the horizontal directions, [CE-23.23, SE=8.74 t(9) = 2.12 p< .001 for LR and CE-14.66, SE=5.16 t(9) = 5.12 p< .001 for RL respectively]. The ANOVAs on Ces confirmed main experiments results, with a significance for the Direction Factor for SP [F (1, 8)=4.4 p< .02] and for the Velocity Factor for the VP [F(1,9)=18.15 p< .000]. For SP, data of Experiment 3 were confronted with data of Experiment 1 and, for VP, data of Experiment 4 were confronted with data of Experiment 2. Given that these experiments served as a control for instantaneous velocity, a new factor (Session, 2 levels) was added in the ANOVAs. For both SP and VP, Factor Session proved not significant: this means that overall there were no difference in participant's performance with or without velocity modulation. This implies that the effect of acceleration on VP errors is due to target's final instantaneous velocity and not to a modulation of it.

These results were also confronted with data obtained from two other experiments (Bastianelli et al., 2004) where velocity was modulated in a sinusoidal way (i.e. the target was firstly accelerating and then decelerating or vice versa) in the central part of the trajectory, with the same initial and final instantaneous velocity used here. In that experiments acceleration had no effect on the SP and VP errors. The fact that here acceleration has an effect on VP error, together with the results of control experiments, is in favour of the hypothesis that forward displacement of VP is based on the local, ongoing motion trajectory (Hubbard and Bharucha, 1988). Furthermore, the results of the whole research (collected till now) supports the idea of a low-level early mechanism, at least for VP dislocation. For SP, in no one of our Experiments acceleration proved significant. This could of course imply that the model proposed by Actis Grosso and Stucchi (2003) is at least not accurate. However, it should be considered that in previous studies (e.g. Hubbard and Motes, 2002) an effect of absolute velocity on the magnitude of SP error was always found: maybe, the range of velocities used in our experiments was too narrow. Next step will be to use a wider range of velocities and to see whether velocity interacts with the presence of a static anchor in the visual scene.

References

Actis-Grosso, R., Stucchi, N.,Vicario, G.B. (1996). On the length of trajectories of moving target dots. In S. C. Masin (Ed.), Fechner Day 1996: *Proceedings of the 12th Annual Meeting of International Society for Psychophysics* (pp. 185-190). Padua, Italy: International Society for Psychophysics.

Actis-Grosso, R., Stucchi, N. (2003). Shifting the start: Backward mislocation of the initial position of a motion. *Journal of Experimental Psychology: Human Perception and performance*, 29, 675-691.

Bastianelli, A., Actis-Grosso, R., Stucchi, N. (2004). Onset and offset localization: the effect of velocity modulation. In A. M. Oliveira, M. Teixera, G. F. Borges, M. J. Ferro (Eds), Fechner Day 2004: *Proceedings of the 20th Annual Meeting of International Society for Psychophysics* (pp. 298-303). Coimbra, Portugal: International Society for Psychophysics.

Fröhlich, F., W. (1923). ber die Messung der Empfindungszeit [On measuring sensation time]. *Zeitschrift f r Sinnesphyfisiologie*, 54, 58-78.

Hubbard, T. L. (1990). Cognitive representation of linear motion: Possible direction and gravity effects in judged displacement. *Memory & Cognition,* 18, 299-309.

Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bullettin and Review*, 2, 322-338.

Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bullettin and Review*, 2, 322-338.

Hubbard, T. L., Bharucha, J. J. (1988). Judged displacements in apparent vertical and horizontal motion. *Perception and Psychophysics*, 44, 211-221.

Hubbard, T. L., Motes, M. A. (2002). Does representational momentum reflect a distortion of length or the endpoint trajectory? *Cognition,* 82, 89-99.

Kerzel, D. (2003a). Attention maintains mental extrapolation of target position: Irrelevant distractors eliminate forward displacement after implied motion. *Cognition*, 88, 109-131.

Kerzel, D. (2003b). Centripetal force draws the eyes, not memory of the target, towards the center. *Journal of Experimental Psychology: Learning, Memory and cognition*, 29, 458-466.

Kerzel, D., Gegenfurtner, K.R. (2004). Spatial distortions and processing latencies in the onset repulsion and Fröhlich effects. *Vision Research*, 44, 577–590.

Kerzel, D., Jordan, J. S., & Müsseler, J. (2001). The role of perception in the mislocalization of the final position of a moving target. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 829-840.

Kirschfeld, K., Kammer, T. (1999). The Fröhlich effect: a consequence of the interaction of visual focal attention and metacontrast. *Vision Research*, 39, 3702-1033.

Müsseler, J., Aschrsleben, G. (1998). Localizing the first position of a moving stimulus: The Fröhlich effect and an attention-shifting explanation. *Perception and Psychophysics*, 60, 683-695.

Müsseler, J., Kerzel, D. (2004). The trial context determines adjusted localization of stimuli: reconciling the Fröhlich and onset repulsion effects. *Vision Research*, 44, 19, 2201-2206.

Nijhawan, R. (1994). Motion extrapolation in catching. *Nature*, 370, 256-257.

Runeson, S. (1974). Constant velocity – Not perceived as such. *Psychological Research*, 37,3-23.

Thornton, I. M. (2002). The onset repulsion effect. *Spatial Vision,* 15 (2), 219-243.

Scholz, W. (1924). Experimentelle Untersuchungen über die phänomenale Grösse von Raumstrecken, die sukzessiv-Darbietung zweier Reize begrentz werden. [Experimental researches on phenomenal size of spatial distances, which are delimited by the successive presentation of two stimuli]. *Psychologische Forschung,* 5,219-272.

Verfaille, K., & d'Ydewalle, G. (1991). Representational momentum and event course anticipation in the perception of implied periodical motions. *Journal of Experimental Psychology: Learing, Memory, and Cognition*, 17, 302-313.

A SIGNAL DETECTION ANALYSIS OF THE EMOTIONAL STROOP EFFECT

Boaz M. BenDavid, Noa Calderon, and Daniel Algom
Department of Psychology, Tel-Aviv University
boazb@post.tau.ac.il

Abstract

The Emotional Stroop Effect (ESE) was tested for accuracy. A severe time limit imposed on reporting produced relatively high rate of errors, documenting the ESE for the first time for accuracy (higher error rates in naming the color of emotional words), and enabling the application of the Theory of Signal Detectability. The results showed that the presence of threat (emotional words) diminished the psychological distance (d’) between the print colors. However, decision (β) remained invariant. These results support an explanation of the ESE as product of a defense mechanism that responds to threat by temporarily appropriating resources away from the threat-irrelevant attributes.

In the "emotional Stroop paradigm," the participants are asked to name the colors in which various words are printed. Naming the color of emotional words is performed slower than that of neutral words. This difference is labeled the "emotional Stroop effect" (ESE) and is theorized (Williams, Mathews, & MacLeod, 1996) to be the product of subliminal processes of word reading interfering with the color naming response. If the participant focuses exclusively on the print colors and ignores the carrier words, then no difference in color naming between emotional and neutral words is expected. The emotional Stroop effect is, therefore, an indication of the failure to pay exclusive attention to print color, and mainly, of the threat-induced disruption of color naming.

The emotional Stroop paradigm is popular in current clinical psychology (see Williams et al., 1996, for review). The effect has been demonstrated with a variety of psychopathological populations, using emotional words related to the respective clinical conditions. There have been relatively few studies investigating the emotional Stroop effect with non-pathological participants (but see, Algom, Chajut & Lev, 2004; McKenna & Sharma, 1995, 2004). Therefore, the ESE is also documented for the normal population.

To date, investigations of the ESE used reaction time (RT) as the indicator of the effect of emotional content on color naming. Accuracy in extant studies was usually close to perfect. In the present study we, for the first time, gauged the ESE for accuracy. Extending the arsenal of ESE measures is important, but the use of accuracy accomplishes more than this limited goal. It enables the application of the Theory of Signal Detectability (TSD) to ESE research and analysis. A range of intriguing questions can be answered.

Does the ESE obtain for accuracy? What, exactly, is impaired under threat? Does threat impair genuine perceptual processing of stimuli in the environment? Alternatively, does threat effect response bias rather than actual stimulus discriminability? Concerning classifications of print colors, do the colors of threat words appear less salient to perception (hence, d' decreases), or does the criterion of reporting them change (hence, β is altered) under threat? These questions were considered in the present pioneering application of the TSD to the ESE.

Therefore, in this experiment emotionally charged and neutral words were presented, in separate blocks, printed in either red or orange font. The participant's task was to decide whether the presented word was printed in red or not in red (i.e., in orange), by pressing the appropriate key. To generate high rates of error, we introduced a time limit of 500 milliseconds for responding (i.e., responses longer than 500 milliseconds were discounted and punished by an unpleasant sound).

Method

Participants

Twenty-eight young men and women from the Department of Psychology, Tel Aviv University, and from the Department of Criminology, Bar-Ilan University, volunteered to take part in the experiment in a partial fulfillment of course requirement. All participants were native Hebrew speakers and had normal or corrected-to-normal visual acuity assessed by self-reports.

Stimuli and apparatus

There were 16 neutral words, names of items of clothing (e.g., hat, shoe), and 16 emotionally charged words associated with terrorism (e.g., terrorist, bomb). Words were balanced in terms of length and frequency of usage within the student population. The stimuli were identical to those used by BenDavid, Levy and Algom (2003), yielding a substantial ESE for RT. Stimulus presentation and measurement were performed by a Macintosh computer with a standard Macintosh keyboard, using PsyScope (version PPC 1.2.5) software. Words were presented in red and orange on a white background in Arial font (size 48), on a 17” color screen.

Procedure.

Participants were tested individually. Each participant performed in two experimental blocks**:** one with 16 neutral words and one with 16 emotional words. The order of the blocks was counterbalanced between participants. Each word appeared 6 times in each of the print colors, red and orange, making for 192 experimental trials per block. Order of trial presentation was random and different for each participant. Each block was preceded by 8 training trials that were later discarded from analysis. Responses were produced by pressing one of two designated keys on a standard keyboard. Key designation was counterbalanced between participants.

The participant was asked to decide, as accurately as possible, whether the print color of the word appearing on the screen was red or not (i.e., orange). Each trial began with the presentation of a mask ("XXXXXXX" in Arial font, size 72, bold) for 500 milliseconds, at the center of the screen. The mask was immediately followed by the presentation of the stimulus. As soon as a response was made, or after a 500 milliseconds time limit, the stimulus was removed for a blank inter-trial interval of 50 milliseconds. The participants were notified that failure to respond within the time limit will be penalized by an unpleasant tone (it was a standard Macintosh error tone) and that the trial will be discarded.

Results and Discussion

Figure 1 gives the results. A glimpse of Panel A shows that the participants were more accurate to discriminate the print colors of neutral words (M= 65.93%) than to discriminate the same colors of emotional words (M= 58.68%). The difference was 7.25% in favor of the former [t(27)=6.66 p<0.001]. Therefore, emotional stimuli effected color naming (= the ESE). For the first time, we recorded an ESE for accuracy (rather than for RT).

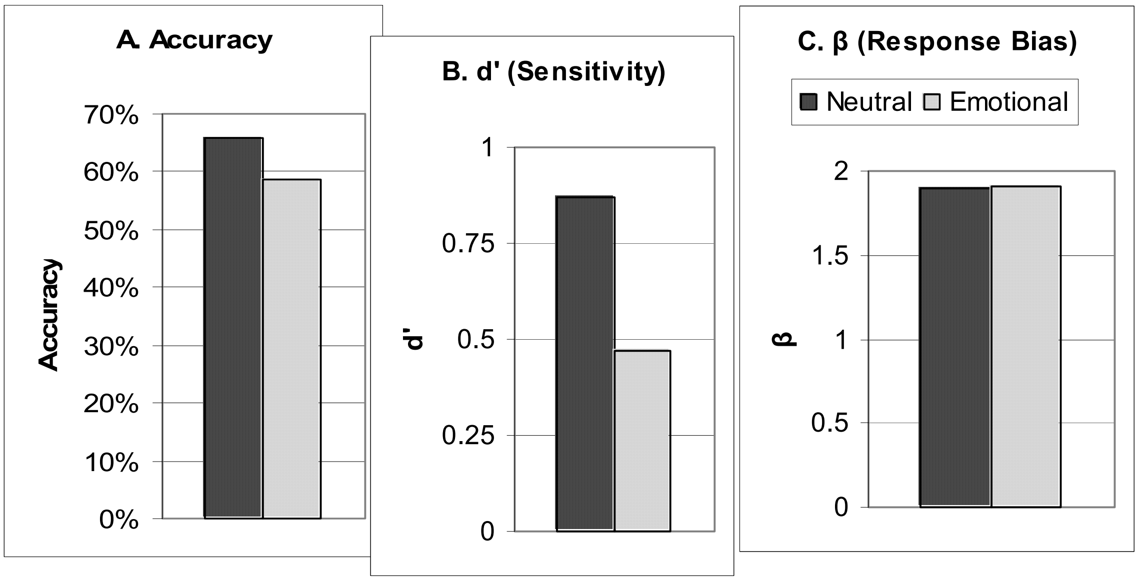


Figure1: A. Accuracy of discriminating print colors for neutral and emotional words; B. Sensitivity measures (d') for discriminating print color of neutral and emotional words; C. Response bias (β) for discriminating print color of neutral and emotional words.

Applying the TSD to these non-speeded data can reveal two possible sources for the disruption: sensitivity, d', or response bias, β. What makes the color naming of emotional words worse than color naming of neutral words? Panels B and C give the answer. Notice in Panel B that the average d', for discriminating print color, of neutral words was larger by 0.4 than the d' for emotional words [0.869, and 0.468 respectively, t(27) = 6.48 p<0.001]. Therefore, the participants were much more sensitive to the difference between the two print colors when the carrier words were neutral than when they were emotional. In contrast, Panel C does not show a difference in β, for discriminating print color, between neutral and emotional words [1.901 and 1.903 respectively, t<1].

These results show that the content of the words exerts a deep sensory effect on the perception of their colors. The presence of threat (emotional words) diminishes the psychological distance (d') between the colors of the carrier words, desensitizing participants to the differences between the colors. Notably, decision making (β) is not altered by threat.

Taken together, the results support an explanation of ESE in terms of a generic inhibitory defense mechanism that responds to threat by temporarily freezing all ongoing activity (Algom et Al., 2004). The presence of threat (emotional words) distorts perception of the threat-irrelevant attributes. The slowdown in the color naming of an emotional word (the emotional Stroop effect) is the product of a preattentive (or an early attentive) inhibitory mechanism that appropriates resources away from the color naming activity. This mechanism is activated by the sheer presence of a threat-inducing stimulus.

References

Algom, D., Chajut, E. & Lev, S. (2003). A rational look at the emotional stroop phenomenon: A generic slowdown, not a stroop effect. *Journal of Experimental Psychology: General, 133,* 323-338.

BenDavid, B., Levy, L. & Algom D. (2003). *The emotional Stroop effect is a generic reaction to threat, not a selective reaction to specific semantic categories.* In: Berglund, B. & Borg, E. (Eds.), Fechner Day 2003. Stockholm: International Society for Psychophysics, 2003, 21-24.

McKenna, F. P., & Sharma, D. (1995). Intrusive Cognitions: An investigation of the emotional Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 21, 1595-1607.

McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 382-392.

Williams, J. M., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120,* 3-24.

Thanks are due to the Department of Criminology, Bar-Ilan University, for their cooperation.

PERCEPTION OF ELLIPTIC MOVEMENTS IN 7-TO-11 YEAR OLD CHILDREN

Christel Bidet-Ildei, David Méary and Jean-Pierre Orliaguet
Laboratory of Psychology and Neurocognition, CNRS UMR 5105 Université Pierre Mendès-France, 38040 Grenoble Cedex 9, France
E-mails: christel.ildei@upmf-grenoble.fr, jean-pierre.orliaguet@upmf-grenoble.fr

Abstract

The present work aims at showing that visual perception is influenced by motor principles. To this end, the capacity of 7-to-11 year old children to perceive elliptic motions was investigated. We tested the respect of the isochrony principle that is the tendency to maintain constant the duration of movement across change in trajectory length. Participants of each group had to adjust the period of a dot depicting an elliptic motion. Six different perimeters (2,94 to 53 cm) were presented on a computer screen. For each subject the relation between the period chosen and the ellipse perimeter was analyzed. Results showed a correspondence with data observed in motor production (Viviani and Schneider 1991): From 7 years-old perceptual adjustments are guide by the isochrony principle. These findings are discussed in the theory of the motor-perceptual link.

The perceptual system exhibits exquisite sensitivity to biological motion. A large body of data has demonstrated that even simplified depiction of human motion, based on point-light display, is sufficient for observers to identify complex actions (Johansson, 1973). One possible account for this tuning is provided by the visual-motor linkage hypothesis. This view assumes that perceivers are guided by motor rules and use their own motor knowledge for the visual perception and identification of biological motion (Méary, Chary, Palluel & Orliaguet, in press; Viviani & Stucchi 1989). Some experiments carried out on children demonstrated that perceptual capabilities were influenced by their own motor competency. For example, Desjardins and colleagues (1997) have shown that the perception of speech is linked to the articulation capacities of children. In the same way, the perceptual anticipation which permits to guess the next component of a motor sequence is influenced by motor anticipation capacity (Louis-Dam, Kandel & Orliaguet, 2000).

The aim of this research is to bring additional evidences that, in children, visual perception is influenced by motor principles. To this end, we studied the velocity adjustments for different sizes of elliptic movements in 7 to 11 years-old children. In production, Viviani & Schneider (1991) have shown, from five-year old children, that elliptic motions are constrained by the isochrony principle. They formalised this principle by a power relation between the movement time and the perimeter of the ellipse with a exponent value close to 0.4.We hypothesised that if perception of elliptic motions is influenced by motor rules we should observe a respect of the isochrony principle i.e. a power relation between the perimeter of the ellipse and the perceptual adjustment with an exponent equal to about 0.4 whatever the age of the subjects.

Method

Participants

Forty-five children separated in three group of age (7, 9 and 11 year old) were recruited as volunteers in different French schools. The standard deviation of the age in each group was ± 4 months. A group of fifteen students of the University Pierre Mendès France was used as control. All subjects were right-handed and had normal or corrected to normal vision. None of them had history of neurological or psychiatric disorders.

Stimuli

The visual stimuli consisted of an elliptic movement (see figure 1). The semi-axis ratio b/a and the eccentricity of the ellipse ∑ were constant and equal to 0.425 and 0.9 respectively. There was 400 pairs of (x,y) coordinates. The major axis of the ellipse *a* was rotated by 45degres. The program permitted to manipulate the trajectory length (L) and the movement period (P) independently. For all period value, the velocity profile of the movement respected the two-third power law observed in the production of movement (Lacquaniti, Terzuolo & Viviani, 1983).

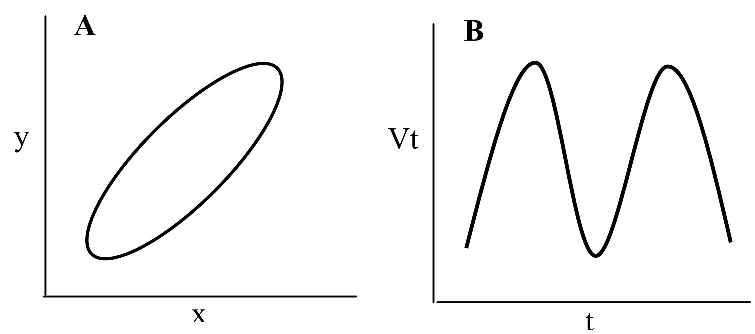


Figure 1: Kinematic characteristic of the stimulus: (A) Elliptic trajectory. (B) Velocity profile of the elliptic movements with Vt corresponded to the tangential velocity and t to the movement time.

Procedure

Participants seated at a distance of 50 cm before a computer screen (17', resolution 1024*768 pixels, sampling rate 85 Hz) in a dimly illuminated room. Each trial consisted on a black spot (ø = 0.4 cm) depicting an elliptic movement on a white background area (22*22 cm). Participants were asked to adjust the period of movement using a computer keyboard in order to find their preferred velocity. The adjustments increase or decrease continuously with step of 25 ms. There was no time limit to do the choice.

The experiment was separated in 6 sessions which corresponded to each size of the ellipse (2,94 ; 5,25 ; 9,36 ; 16,69 ; 29,74 ; 53 cm). The subject had to adjust 6 times the period of movement resulting 36 trials altogether. The session order, the trials in a session, and the initial period value of each trial (pi) were randomized. There was a break between two sessions.

Results and Discussion

To facilitate the comparison with the production task, the results were analysed by using the same procedure that Viviani & Schneider (1991). Therefore, the link existing between the perimeter of the ellipse (Pe) and the final period (Pf) chosen by the participants was approximated by the $Pf = P0*Pe^{\gamma}$ where γ is the exponent of the function and P0 a constant which depends of each subject. These two parameters permit respectively to evaluate the degree of isochrony and the basic tempo of each subject. They were estimated for each participant by a linear regression between the log of the

perimeter and the log of the final period (see figure 2A). The linear correlation coefficient permits to measure the coherence of results and to evaluate the power approximation of the data. The slope is an estimator of the exponent of the power function (γ) and the intercept parameter (log Po) corresponds after transformation[1] to the tempo of the subject.
The mean of the linear correlation coefficient, the slope, and the intercept (log Po) were calculated and statistically evaluated with an ANOVA with age as between factor.
Because the correlation coefficients did not follow a normal law we used the Fisher hyperbolic tangent transform (Kendall & Stuart, 1979).

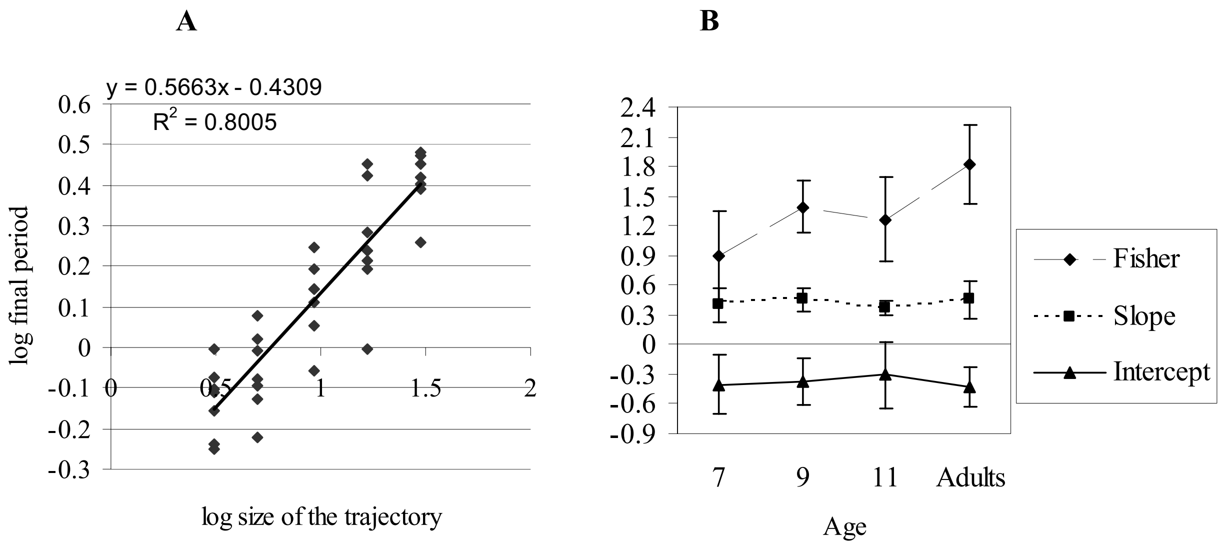


Figure 2: (A) Example of data analysis of one subject. (B) Mean and standard deviation for the fisher, the slope and the intercept according to the age.

Concerning the correlation coefficient the statistical analysis revealed a significant effect of the age ($F_{3,55}$ =12.74, p<0.01). In particular, we observed an increasing between 7 (0,71) and 9 years-old (0,88), a stabilisation between 9 (0,88) and 11 (0,85) years-old and a new increasing between 11 (0,85) years-old and adults (0,95). This result indicates that the variability (intra) is more important at 7 years-old and tends to decrease with age. We can notice that the linear coefficients are very important whatever the age which confirm the good approximation of the power function to evaluate the link between the final period chosen by the subject and the size of the trajectory.

Concerning the slope, there was no significant difference with the age ($F_{3,55}$ =1.12, p=0,35). The exponent of the power regression did not move from 7 year-old. Moreover, the mean value of the exponent is 0.42 (SD=0.15) which is not different of the standard value 0.4 (t_{59}=0.91 p=0.37).
Therefore, perception of elliptic movements is influenced by the isochrony principle. The stable value of γ observed very early in the development can be considered as a constant defining the link between the perimeter and the time movement for elliptic movements.

Concerning the intercept, there was no significant difference with the age ($F_{3,55}$ =0.48, p=0.70). This

[1] The transformation consists to apply an exponent 10 at the intercept value.

intercept corresponded to a mean basic tempo of 0.42 (s). We can notice that this result is different that those obtained in production (Viviani & Schneider 1991): The values obtained here are lower and more stable whatever the age. They could be explained by the task constraints which are different in the two experiments. In production of movement, participants had to determine in the same time the variability and the velocity of their movements. We can suppose that the motor expertise can influence this relation. On the contrary, in the perception task there was no variability of movement because stimuli are entirely defined to respect exactly a given trajectory of movement. Consequently, the adjustment of the velocity is only dependant on the same defined spatial constraint. We can hypothesis that the perceptual tempo of the subjects was inferior because subjects have adjusted the velocity of the movement to respect the form of the trajectory.

The most important finding is that perceptual judgements of elliptic motion are similar to those obtained in production (Viviani & Schneider, 1991). These findings bring an additional argument in favour of the influence of motor rules in visual perception in children. They are in accordance with neuroimaging and neuropsychological works carried out on adults. For example, clinical observations showed that specific motor pathology like dysgraphia entailed corresponding perceptual difficulties (Chary et al. 2004). In addition, cerebral imaging studies reveal that brain areas activated for the production and perception of human motion are largely overlapping (Rizzolatti et al, 1996; Decety et al, 1997; Hari et al. 1998; Nishitani & Hari 2000).

References

Booth, A. E., Bertenthal, B. I., Pinto J. (2002). Perception of the symmetrical patterning of human gait by infants, *Developmental Psychology,* 38, 554-563.

Chary, C., Méary, D., David, D., Moreau, O. & Kandel, S. (2004). Influence of motor disorders on visual perception of human movements in a case of peripheral dysgraphia, *NeuroCase*, vol. 10, 3, 223-232.

Decety J., Grèsez J., Costes N., Perani D., Jeannerod M., Procyck E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy, *Brain,* 120 1763-1777.

Desjardins, R.N., Rogers, J., Werker, J.F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks, *Journal of Experimental Child Psychology*, 66, (1), 85-110.

Hari R., Forss N., Avikainen S., Kirveskari E., Salenius S., Rizolatti G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study, *Neurobiology,* 95**,** (25), 15061-15065.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14, 201-211.

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics* (vol.2). London: Charles Griffin.

Lacquaniti, F., Terzuolo, C. A., & Viviani, P. (1983). The law relating kinematic and figural aspects of drawing movements. *Acta Psychologica*, 54, 115-130.

Louis-Dam, A., Kandel, S. & Orliaguet, J.P. (2000). Perception visuelle des mouvements humains : anticipation motrice et anticipation perceptive. *Psychologie Française*, 45, 4, 333-342.

Méary, D., Chary, C., Palluel, R. & Orliaguet, J-P. (in press) Visual perception of writing and pointing movements, *Perception.*

Nishitani, N., Hari, R. (2000). Temporal dynamics of cortical representation for action, *Neurobiology,* 97**,** (2), 913-918.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., et al. (1996). Localisation of grasp representation in humans by PET: observation vs execution, *Experimental Brain Results,* 111**,** 246-252.

Viviani, P. & Schneider, R. (1991). A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology*, 17, 1, 198-218.

Viviani, P. & Stucchi, N. (1989). The effect of movement velocity on form perception : Geometric illusions in dynamic displays. *Perception & Psychophysics,* 46, 266-274.

PERCEPTION OF ELLIPTIC BIOLOGICAL MOTION

Christel Bidet-Ildei [1,2], Jean-Pierre Orliaguet [2], Alexander Sokolov[3], and Marina Pavlova [1]

[1] Department of Paediatric Neurology and Child Development,
Children's Hospital, University of Tübingen, Hoppe-Seyler-Str. 1, D 72076 Tübingen, Germany
[2] Laboratory of Psychology and Neurocognition,
CNRS UMR 5105 Université Pierre Mendès-France, 38040 Grenoble Cedex 9, France
[3] Center for Neuroscience and Learning and Department of Psychiatry,
University of Ulm Medical School, Leimgrubenweg 12-14, D 89075 Ulm, Germany
e-mails: christel.ildei@upmf-grenoble.fr, marina.pavlova@uni-tuebingen.de

Abstract

We tested the capability of the mature visual system for discrimination between types of elliptic biological motion on the basis of event kinematics. Healthy adult volunteers were presented with point-light displays depicting elliptic motion when only a single dot, a moving point-light arm, or a whole point-light human figure was visible. The displays were created in accordance with the two-thirds power kinematic law (*natural* motion), whereas the control displays violate this principle (*unnatural* motion). On each trial, participants have to judge whether the display represents natural or unnatural motion. The findings indicate that adults are highly sensitive to violation of the two-thirds power kinematic law. Notably, participants can discriminate between natural and unnatural motions without recognizing the stimuli that suggests implicit using of kinematic information. Most intriguing, event recognition seems to diminish the capacity to judge whether event kinematics is unnatural. We discuss possible ways for a cross talk between perception and production of biological movement.

More than three decades ago, Gunnar Johansson introduced point-light displays depicting different types of human locomotion (Johansson, 1973). Since that time, one of the most robust and replicable findings in the field is that the mature visual system is exquisitely tuned to biological motion produced by living organisms, and a number of studies attempt to clarify this extraordinary tuning. One possible account for the high sensitivity to biological motion is provided by the visual-motor linkage hypothesis. This view assumes that perceivers use their own motor knowledge for the visual perception and interpretation of biological movement. In accord with this, it is reported that observers better recognize a point-light display representing their own movement for which they have implicit motor programs than movements of their friends with which they have only visual experience (Loula, Prasad, Harber, & Shiffrar, 2005). Human infants also exhibit a kind of complementary link in the development of the motor skills and abilities to perceive biological movement (Booth, Bertenthal, & Pinto, 2002).

Psychophysical data indicate that visual processing of bodily movements is substantially affected by the biomechanical properties inherent for these movements (e.g., Shiffrar & Freyd 1993; Pavlova & Sokolov 2000). In particular, spatial and kinematic constraints are of immense value for successful recognition. Display inversion in the image plane dramatically impedes veridical perception of a point-light walker, presumably because of altered information about dynamic gravity-inertial force vectors (Pavlova & Sokolov, 2003; Shipley, 2003). This inversion effect is diminished for biomechanically impossible body poses (Reed, Stone, Bozova, & Tanaka, 2003).

The role of kinematic constraints in biological motion processing is reflected by the two-thirds power law pointing to a specific link between the tangential velocity and the trajectory path

(Lacquaniti, Terzuolo, & Viviani, 1983). This law is defined by the equation: $V(t)=K(t)*\{R(t)/1+\alpha R(t)\}^{1-\beta}$ where V(t) is the tangential velocity, R(t) is the radius of curvature of the trajectory in time t, K is a velocity gain factor dependent on the movement tempo, α is a coefficient dependent on the inflection point on the trajectory, and β is an exponent which has a value very close to 2/3 from puberty (Viviani & Schneider, 1991). Motion corresponding to the two-thirds power law can cause rather strong perceptual illusions (Viviani & Stucchi, 1989, 1992). For example, when a dot moves along a circular trajectory with an instantaneous velocity corresponding to the two-thirds power law (with kinematic characteristics corresponding to an elliptic path) subjects perceive an elliptic trajectory instead of a circular one.

In the present work, we investigated the capability of the mature visual system for discrimination between elliptic motions on the basis of their kinematics that either corresponds to the two-thirds power kinematic law (*natural*) or violates it (*unnatural* motion). In particular, we addressed the following issues: (i) whether the visual system is tuned to specific changes in the distribution of velocity, and (ii) whether recognition of events is an obligatory prerequisite for discrimination between *natural* and *unnatural* event kinematics. With this purpose in mind, we presented healthy adult observers with point-light displays depicting elliptic motion when only a single dot, a moving point-light arm, or a whole point-light human figure was visible. In other words, the displays varied in the amount of contextual information that might promote event recognition.

Method

Participants

Twelve right-handed adults (7 males, 5 females, aged between 19 and 43, mean age 27.5 (±5.47) years), with normal or corrected-to-normal vision participated in the study. Nobody had a history of neurological or psychiatric disorders. Participants did not have any previous experience with biological motion stimuli, and were unfamiliar with the purpose of the study. They were recruited as volunteers among the international community of the University Hospital of Tübingen, Germany.

Stimuli

We used three types of point-light displays depicting an elliptic motion (Figure 1): a point-light human figure, a point-light arm, and a single dot. The first type of stimuli depicted an elliptic movement produced by the left rigid arm of a human figure represented by 14 dots placed on the main joints (shoulders, elbows, hands, hips, knees, and feet), the pelvis and the head. The whole figure except for the left shoulder, elbow, and hand was static. The elliptic arm movement was represented by 9 dots placed on the head, the shoulders, the elbows, the wrists and the index finger of an actor viewed from above. Elliptic motion spontaneously produced by the left hand of the actor holding a pencil was filmed by the 3-D ultrasound Zebris system that allows for recording the spatial and temporal parameters of movement (spatial resolution; 0.2 mm; temporal resolution, 200 Hz). The third type of stimuli depicted an elliptic motion of a single dot placed on the index finger of the left arm. The single-dot and human-figure displays were computer-generated. All displays consisted of bright dots moving against a dark background (Figure 1). The elliptic motion of all three types of stimuli followed the two-thirds power law with ß of 2/3 that resulted in an exponent 1-ß equal to 1/3. We consider this condition as a natural motion. For each type of stimuli, control displays were created by violating this kinematic law. Namely, in the equation given above we changed the exponent ß to 4/3 (1-ß=-1/3) to obtain an inverse velocity profile that gave an impression of abrupt alterations of accelerations and decelerations. Dots in all displays moved along the elliptic trajectory with a mean velocity of 60 cycles/min, one cycle corresponds to 40 frames with frame duration of 25 msec. The perimeter of the elliptic form was 3.87, 10.41, and 19.31 cm for the human-figure, the arm, and the single-dot display, respectively.

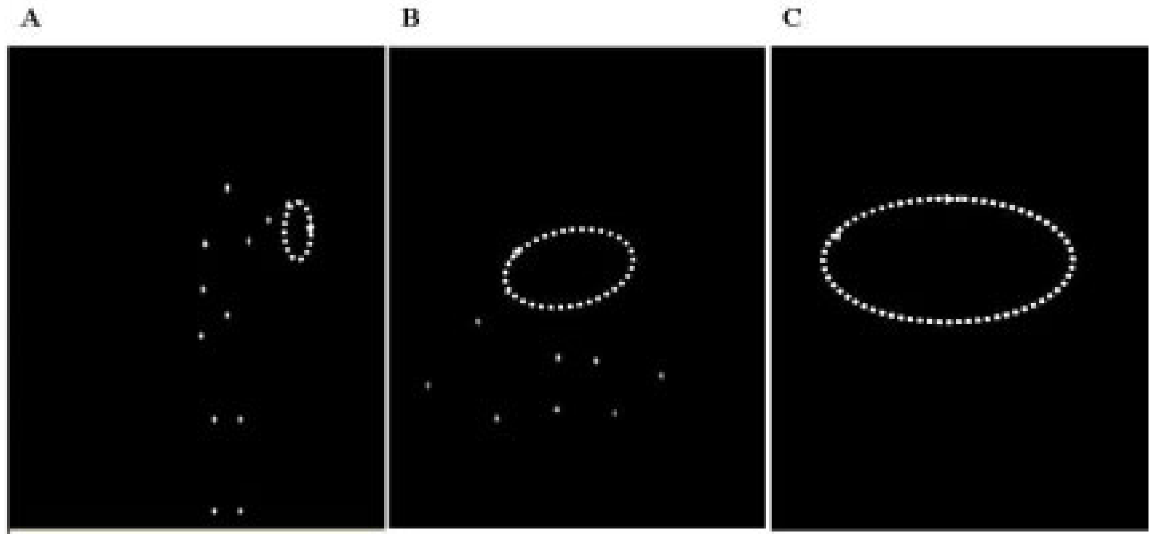


Figure 1. Static representation of the point-light displays: (A) a human figure, front view; (B) an arm, view from above; and (C) a single dot, front view. Light arrows indicate directions of motion.

Procedure

Participants were seated at a distance of 70 cm in a front of a computer screen (800*600 pixels, 60 Hz) in a dimly illuminated room. In each run, fifty trials were presented for each display type and condition (natural/unnatural) in a randomised order. On each trial, participants judged as quickly and accurately as possible whether the display depicts natural or unnatural biological movement. Participants pressed two response keys on a panel indicating either natural or unnatural motion. They used the right hand for their responses. The assignment of response keys (left-right for *natural* and *unnatural* motion) was counter-balanced across participants. After the whole experimental run was completed, the participants were asked to briefly describe what they saw and to indicate any stimulus interpretation they might have had. The whole experimental session took about 20-25 minutes.

Results

Discrimination

For each subject and display type, a d' value, a standard measure of sensitivity in signal detection theory, was computed. For all displays, d' values were higher than 2.5 (2.71, 4.71, and 2.86, for human figure, arm, and single dot, respectively) that indicates extremely good capability for discrimination. A one-way ANOVA with factor display type (single dot/arm/human figure) performed on individual d' values revealed a significant difference in sensitivity between the types of stimuli ($F_{2,22}$=8,38, $p<0.01$). Pair-wise comparisons indicated that the sensitivity for the movement of a point-light arm is higher than for the single dot (t_{11}=3.39, $p<0.01$) and for the point-light human figure (t_{11}=3.37). Despite the differences in the extent of motion path, no difference in sensitivity was found between the single-dot display and the human-figure display (t_{11}=0.32, n.s.).

Overall, participants gave many more hits to all types of displays under the natural (90.94%) than under the unnatural condition (78.33%). This indicates that when event kinematics violates the two-thirds power law, it is much more difficult to judge whether an elliptic motion is natural. For the arm and single-dot displays, the percentage of hits was above chance both under the natural and unnatural conditions (t_{11}=42.60; t_{11}=12.80, for the natural, and t_{11}=22.36, t_{11}=8.42 p<0.001, for the unnatural condition, respectively). For the human-figure display, the percentage of correct responses was above chance only under the natural condition (t_{11}=26.78, p<0.001).

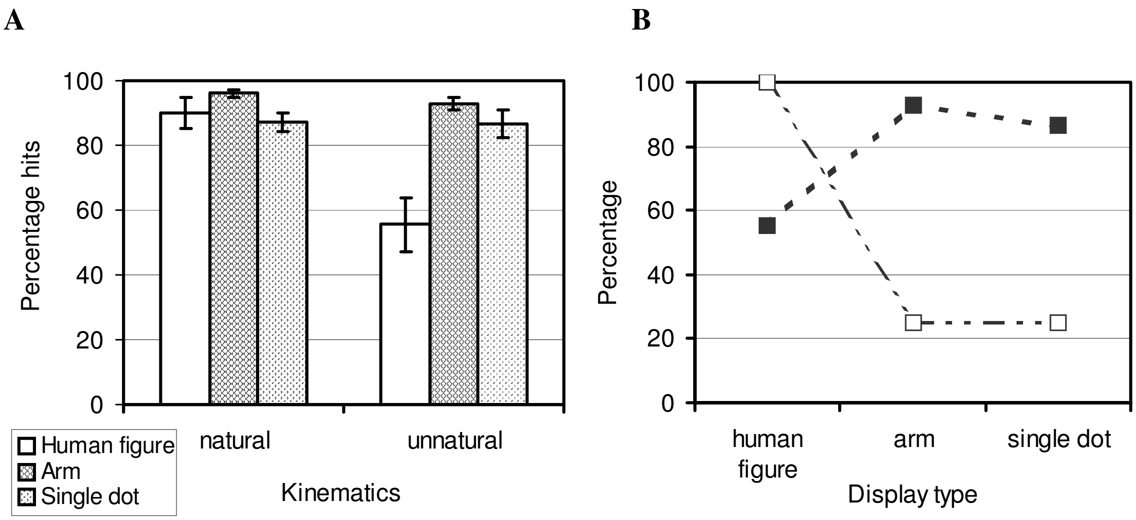


Figure 2. (A) Percentage of hits for different display types under the natural and unnatural conditions. Vertical bars represent ±SE. (B) Comparison between percentages of hits in the discrimination task (filled squares) and correct display recognition (open squares) under the unnatural condition.

Recognition

All participants recognized the display depicting a human point-light figure as a human being producing an elliptic motion by his/her hand. Only three out of 12 participants recognized an elliptic motion in the single-dot and arm displays. The majority of participants interpreted these displays as representing motion of dots without any event identification or naming. One participant interpreted the point-light arm display as somebody who is swimming, and one participant reported the elliptic-like motion of a leg. One participant interpreted the single-dot display as a flying mosquito. As can be seen from Figure 2B, the percentage of hits in unnatural conditions for all display types is inversely related to display recognition: the most recognizable display of a human figure elicits the lowest number of hits.

Discussion

The findings clearly indicate that healthy adults are highly tuned to natural and unnatural biological motions, and they can make their judgements solely on the basis of velocity distribution along motion trajectories. Notably, participants make less erroneous responses when perceiving elliptic motion that corresponds to the principle of biological movement production. In general, these findings appear to corroborate psychophysical, neuroimaging and neuropsychological work revealing the link between production and perception of human movements. For example, when adults and children are choosing the next step in a perceived sequence of handwriting or reaching movements, they rely upon the motor anticipation rule that is determined by production of motor sequences (Louis-Dam, Kandel, & Orliaguet, 2000). Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) studies reveal that brain areas activated for the production and perception of biological motion are largely overlapping (Rizzolatti et al., 1996; Decety et al., 1997; Hari, et al., 1998; Nishitani & Hari 2000). For example, Broca's region and the precentral motor cortex are activated both during execution of hand actions and during observation of similar actions performed by other individuals (Rizzolatti et al., 1996). Neuropsychological work with patients also indicates a close link between production and perception of motion. For example, a patient with dysgraphia exhibits correspondences between difficulties in the production and perception of handwriting motions (Chary et al., 2004).

The other important outcome of this work is that discriminability of natural and unnatural kinematics is not necessarily linked to event recognition. The natural and unnatural kinematics of two out of three displays was highly discriminable, but almost all participants failed in their spontaneous recognition. One possible explanation for this finding consists in the implicit use of motor knowledge in visual processing of biological movement. In accord with this, Louis-Dam with colleagues (2000) have reported that in a handwriting anticipation task perceivers are unable to explain or give pertinent indications guiding their perceptual anticipation. An additional argument in favour of the implicit use of motor rules in biological motion processing comes from the notion about the existence of two types of representations (Jeannerod & Jacobs, 2005). One of them, pragmatic representation mode, corresponds to procedural capacities of the perceivers, and contains the representation of his/her motor abilities. This type of representation is very difficult to verbalize because it is based on implicit knowledge about how to produce movements. The other, semantic mode, contains explicit representations that are easy to verbalize.

The most intriguing finding is that event recognition seems to diminish the capacity to judge whether event kinematics is unnatural: when the display kinematics violates the two-thirds power law, the most recognizable whole body display elicits the lowest percentage of hits. In other words, when a display is easily recognizable as a human figure, observers seem to be biased toward interpretations of this display as representing natural motion. This appears to be in agreement with the findings that the perceived identity and previous experience associated with this identity can influence how that figure is perceived to move. For example, Brosgole and Whalen (1967) reported that an inducing object was more effective when it induced an airplane to move forward rather than backwards. The other example might be that the recognition of point-light displays substantially diminishes the apparent-facing effect caused by reverse transformation of biological motion (Pavlova, Krägeloh-Mann, Birbaumer, & Sokolov, 2002). There is a bias to perceive forward locomotion, at the expense of misinterpreting the underlying form in time-reversed biological motion films. Overall, it appears that knowledge about object nature helps one to tolerate visual distortions, but it also can lead to erroneous perceptual interpretations. Although it is known that top-down influences can profoundly modulate visual sensitivity in static displays, there is only limited evidence indicating that dynamic displays can also be affected. Future research, therefore, is required for investigation of this phenomenon.

Acknowledgements

We thank David Méary for assistance in displays creation, Jürgen Dax for creating the program for stimuli presentation, John C. Baird for valuable advice on the manuscript, and Ingeborg Krägeloh-Mann for general support. C.I. was a visiting PhD student in M.P.'s lab, her stay was supported by the "Aires Culturelles" Program of the French Research Ministry. The paper presentation at the Annual Meeting of the International Society for Psychophysics, Fechner Day 2005, is supported by a travel grant to C.I. from the International Society for Psychophysics.

References

Booth, A. E., Bertenthal, B. I., & Pinto, J. (2002). Perception of the symmetrical patterning of human gait by infants. *Developmental Psychology,* 38, 554-563.

Brosgole, L., & Whalen, P. M. (1967). The effect of meaning on the allocation of visually induced motion. *Perception & Psychophysics*, 2, 275-277.

Chary, C., Méary, D., Orliaguet, J.-P., David, D., Moreau, O., & Kandel, S. (2004). Influence of motor disorders on visual perception of human movements in a case of peripheral dysgraphia. *Neurocase,* 10, 223-232.

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyck, E., Grassi, F., & Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain,* 120, 1763-1777.

Jeannerod, M., & Jacob, P. (2005). Visual cognition: a new look at the two-visual systems model. *Neuropsychologia,* 43, 301-312.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysic*s, 14, 201-211.

Hari, R., Forss, N., Avikainen, S., Kirveskari, E., Salenius, S., & Rizzolatti, G. (1998). Activation of human primary motor cortex during action observation: a neuromagnetic study. *Neurobiology,* 95, 15061-15065.

Lacquaniti, F., Terzuolo, C. A., & Viviani, P. (1983). The law relating kinematic and figural aspects of drawing movement. *Acta Psychologica (Amsterdam),* 54 ,115-130.

Louis-Dam, A., Kandel, S., & Orliaguet, J-P, (2000). Anticipation motrice et anticipation perceptive. *Psychologie Française,* 45, 333-342.

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Psychophysics,* 31, 210-220.

Nishitani, N., & Hari, R. (2000). Temporal dynamics of cortical representation for action. *Neurobiology,* 97, 913-918.

Pavlova, M., Krägeloh-Mann, I., Birbaumer, N., Sokolov, A. (2002). Biological motion shown backwards: the apparent facing effect. *Perception,* 31, 435-443.

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics,* 62, 888-899.

Pavlova, M., & Sokolov, A. (2003). Prior knowledge about display inversion in biological motion perception. *Perception,* 32, 937–946.

Reed, C., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science,* 14, 302-308.

Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., Fazio, F. (1996). Localisation of grasp representation in humans by PET: 1. Observation versus execution. *Experimental Bran Results,* 111, 246-252.

Shipley, T. F. (2003). The effect of object and event orientation on perception of biological motion. *Psychological Science,* 14, 377-380.

Shiffrar, M., & Freyd, J. J. (1993). Timing and apparent motion path choice with human body photographs. *Psychological Science,* 4, 379-384.

Viviani, P., & Schneider, R. (1991). A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology: Human Perception and Performance,* 17, 198-298.

Viviani, P., & Stucchi, N. (1989). The effect on movement velocity on form perception: geometric illusions in dynamic displays. *Perception & Psychophysics,* 46, 266-274.

Viviani, P., & Stucchi, N. (1992). Biological motion looks uniform: Evidence of motor-perceptual interactions. *Journal of Experimental Psychology: Human Perception and Performanc*e, 18, 603-623.

CHEATER RECOGNITION AND THE AVOIDANCE OF EXPLOITATION IN SOCIAL EXCHANGE

Dan Boyll, Sonya Sheffert, Steven Colarelli, and Bryan Gibson
Department of Psychology, Central Michigan University, 101 Sloan Hall, Mt. Pleasant, MI
boyll1dl@cmich.edu

Abstract

An understanding of natural selection as a gene-level phenomenon dictates that we understand group-directed behavior, such as cooperation, as having evolved because it is individually advantageous. For cooperation to be adaptive for the cooperator, however, requires that it be performed judiciously; that is, "cheaters," who will selfishly exploit cooperative gestures, must be avoided. While previous research has demonstrated that cheaters' faces are recognized more frequently than non-cheaters' faces (Mealey et al., 1996; Oda, 1997; Yamagishi et al., 2003), such a bias fails to address this crucial task of cheater avoidance. The current study therefore employed mixed-motive games (against simulated "cheaters" and "non-cheaters") as a means of assessing whether the "cheater recognition bias" could be more parsimoniously understood as a by-product of a more obviously adaptive "cheater avoidance bias." Some evidence was found to support this hypothesis, and implications and important directions for future research were discussed.

Beginning with Hamilton's (1964/1980) introduction of a genetically inclusive concept of fitness, the modern, prevailing view regarding natural selection is that it is a process which takes place at the level of the gene (see also Dawkins, 1976). This "gene's eye view" of selection dictates that we understand cooperation, in which two or more individuals pool their efforts toward a mutually beneficial outcome, as having evolved because it has been individually advantageous (i.e., adaptive for an individual's own genes), even when it takes place between genetically unrelated organisms. However, an important caveat of cooperation is that the potential generally exists for one party to "cheat" the other by exploiting his or her cooperative gesture. For example, if you and I have agreed to cooperate by harvesting a field of apple trees and then splitting the bounty, I could exploit you by falsely claiming that my half of the field yielded substantially fewer apples than it actually did. Assuming you were truthful in splitting your apples with me, I would walk away with a net benefit greater than yours, since I would have half of the apples from your trees and more than half of them from mine. Thus, in light of the potential for this greater net benefit, honest cooperation can be understood to involve an opportunity cost.

Maynard Smith (1982) has made this caveat of cooperation precise via the use of mixed-motive games designed to mimic real-world cooperation decisions. Of these games, the "Prisoner's Dilemma" (PD) is the most widely known. The unifying feature of all mixed-motive games is that players must decide between an option that would render the best individual outcome and an option that would render the best mutual outcome. An example of a PD payout matrix is shown below.

S < P < R < T (Example: S = 0, P = 1, R = 3, T = 5)		PLAYER 2	
		Cooperate	*Defect*
PLAYER 1	*Cooperate*	**R, R** (3, 3)	**S, T** (0, 5)
	Defect	**T, S** (5, 0)	**P, P** (1, 1)

*Note: Player 1's payout is listed first.

As Maynard Smith has demonstrated mathematically using computer models of these mixed-motive games, a population including both indiscriminate cooperators and indiscriminate defectors will eventually become dominated by the defectors.

In light of the above evidence, the question becomes: why, then, is cooperation so readily apparent in the human species? The answer lies in our ability to be judicious with our cooperative gestures and in the fact that we repeatedly interact with the same individuals over long periods of time. The most successful PD program in a series of computer tournaments conducted by Axelrod (Axelrod & Hamilton, 1981; Axelrod, 1984), called "tit-for-tat," would always attempt to cooperate in its first game with any other program, but would then mimic that program's most recent decision. In doing so, tit-for-tat effectively avoided being continually exploited by "cheater" programs that had defected against it in the past, at least until those programs switched to a more cooperative strategy. Thus, tit-for-tat was "nice" (always cooperated on the first trial), "retaliatory" (responded in kind to defection), and "forgiving" (resumed cooperating when its opponent did). The analog to human cooperation, of course, is that cooperators could eventually dominate the population as long as they are not so naïve as to be continually exploited by habitual cheaters.

One weakness of tit-for-tat is that it must first be exploited in order to learn who is and who is not a cheater; this will be hereon referred to as tit-for-tat's "interaction requirement." However, a recent line of evidence would seem to suggest that humans could have evolved cooperation strategies that are more sophisticated than tit-for-tat. In recognition studies, several researchers have observed that their subjects were particularly good at recognizing the faces of "cheaters." However, these studies' operational definitions of "cheaters" differed importantly, with some classifying them as such by affixing text labels to their facial photographs during the "study" phase (Mealey, Daood, & Krage, 1996; Oda, 1997) and another classifying them as such based on previous PD decisions and without the use of labels (Yamagishi, Tanida, Mashima, Shimoma, & Kanazawa, 2003).

Since the mere recognition of a cheater would not be adaptive unless it somehow fostered the withholding of cooperative gestures toward that cheater, the current study sought to determine whether modern humans might display a "cheater avoidance bias" by more frequently withholding cooperative gestures from cheaters than from non-cheaters. To this end, the "new" vs. "old" judgments typically made during the "test" phase of recognition studies were replaced by simulated PD games where subjects played "against" the facial photographs presented to them. If evidence for such a bias were observed, it would suggest that modern humans are capable of a strategy of cooperation that is both more sophisticated and, since it avoids any interaction requirement, more adaptive than tit-for-tat.

Further, in light of previous research's divergent operational definitions of "cheaters," the current study varied its method of "cheater" classification, identifying them as such with text labels in one condition and without them in another. This manipulation was also intended to more precisely delineate the nature of any observed cheater avoidance bias; that is, it was intended to answer the question, "Can humans only avoid cheaters when they have prior, explicit knowledge regarding who is and who is not a cheater, or can they make this inference merely by perceiving the faces of those with whom they interact?" Either form would represent a cooperation strategy more sophisticated and more adaptive than tit-for-tat; however, in light of humankind's capacity for observing and gathering reputation information about others, and in light of our dismal skills in detecting deception when we have no explicit cues (Ekman, 1996), it was predicted that cheaters would be more frequently avoided than non-cheaters only when subjects were given prior explicit knowledge via labeling during the initial, "study" phase.

Method

Collection of Stimulus Targets

For the purpose of collecting stimulus targets, a preliminary experiment was first conducted in which subjects played PD games for real money against anonymous partners. In all, 140 students from Central Michigan University (96 females, 44 males) participated in the preliminary experiment in exchange for extra credit in their psychology courses and the opportunity to win a small amount of money (up to $3.00). To ensure the anonymity of PD partners, at least 3 subjects were required for each PD game. Following the PD game, subjects completed a post-game questionnaire (adapted from Yamagishi et al., 2003) and were asked to provide, in an adjacent studio, video footage of themselves uttering a standard phrase ("The tide was out and the sun shone down on the white sand of the beach"; Yarmey, Yarmey, Yarmey, 1994). Video stills were then extracted from non-speaking portions of these recordings; however, due to the lack of male PD players who had defected and who were also without distinguishing features such as facial hair or baseball caps, only video stills of female PD players were considered as potential stimulus targets. Facial attractiveness and expressiveness were controlled for by collecting ratings along these dimensions from a group of independent judges (3 males, 2 females) and eliminating from consideration any video stills which were rated two standard deviations or more away from either mean. Of the remaining 56 video stills, 14 depicted subjects who had defected in the PD game; these were therefore classified as "cheater" stimulus targets. Based on a "cooperation index" calculated from subjects' post-game questionnaire responses, another group of 14 video stills were classified as "cooperator" stimulus targets. These two groups of stimulus targets were not significantly different with regard to subjects' age, t (26) = -0.434, p > .05; average attractiveness rating, t (26) = 0.116, p > .05; or average expressiveness rating, t (26) = 0.982, p > .05.

In addition to these stimulus targets, 18 coat-of-arms images were selected to serve as stimuli for a distractor recognition task. These were collected by performing a "Google" (http://www.google.com) image search for the keywords "coat of arms" and then saving selected results as JPEG computer image files. The only qualifying criteria for the selection of a coat-of-arms image were the absence of text (such as a family name) and a reasonable dissimilarity from other targets already selected.

Participants

Fifty-seven students from Central Michigan University (41 females, 16 males) participated in the experiment in exchange for extra credit in their psychology courses and the opportunity to win a small amount of money (up to $5.00). The only qualifying criterion was that subjects had not participated in the preliminary PD game.

Materials

In order to randomize presentation order, the stimulus targets were ordered into six different presentation lists using a table of random numbers. Three of these lists were then randomly selected to serve as "exposure" lists (shown to subjects as the first presentation list, analogous to the "study" phase of a recognition experiment) while the other three were designated as "game" lists (shown to subjects as the second presentation list, analogous to the "test" phase of a recognition experiment). Each exposure list was then paired, at random, with a game list. Microsoft PowerPoint slideshows were then created by pasting, on every other slide, a stimulus target or distractor target, such that three complete lists were created, all of which had the same progression: stimulus target exposure list, then distractor study list, then distractor test list, and finally stimulus target game list. Instructional slides were also included at the beginning of each list, including several devoted to the nature of the PD. Subjects were instructed that they would be playing several series of simulated PD games, with 14 games in each series; they could pass at any time so as to save game opportunities for later "opponents." Subjects were also told that their opponents' PD decisions could change between series. Each slideshow was automated so that all slides were shown for a uniform amount of time (4 s for exposure and study lists, 6 s for test and game lists) with a uniform 1 s interstimulus interval. Finally, a series of 80 blank slides were added to the end of each slideshow.

The slideshows thus created were saved and used as "Unlabeled" (L-) lists. To create "Labeled" (L+) lists, text boxes were added below stimulus targets in the exposure lists. For cheater stimulus targets, the text boxes read "I confess," and for non-cheater stimulus targets, the text boxes read "I keep silent"; these labels correspond to the classic PD example (in which arrested partners-in-crime are interrogated separately), which was used to instruct subjects on the nature of the PD.

Slideshows were presented to subjects on individual personal computers in a small, 4-computer laboratory. They marked their recognition judgments (for the distractor task) and PD decisions on a numbered response sheet.

Procedure

Subjects were randomly assigned to participate in either the L- or L+ condition. Every effort was made to balance the number of male and female subjects in each condition.

Upon arrival, each subject took a seat at one of the computers and signed a consent form. When all subjects arrived, the experimenter gave a brief summary of the subjects' task and paused for questions. If subjects felt they understood the task, they began the slideshow by pressing the F5 button on their keyboards. At the end of the game list, subjects handed their response sheets to the experimenter and received a new sheet on which to record their PD decisions for the next series of games. However, because only blank white slides remained in all slideshows, the next series never began. The experimenter soon "noticed" this problem, announced that the slideshows were "incomplete," and offered to grant the subjects normal credit with their cash rewards based on their performance in the intact series. No subject took issue with this arrangement. The experimenter then tabulated each subject's score by comparing the subject's decision for each simulated game with against a key indicating the order to "cooperate" and "defect" decisions for each list. Subjects then received their payouts and were thanked and dismissed.

Results and Discussion

Subjects' PD Decisions Against Cheater and Non-Cheater Stimulus Targets

Critical dependent measures included cooperation rate (the proportion of the time that subjects cooperated, or "kept silent"), defection rate (the proportion of the time that subjects defected, or "confessed"), and pass rate (the proportion of the time that subjects decided not to play against the presented opponent). A fourth statistic, avoidance rate, was computed simply as the proportion of the time that subjects avoided cooperation; thus, any given subject's avoidance rate was equal to the sum of their defection and pass rates. Each of these four statistics was computed both across all stimulus targets as well as separately for cheater and non-cheater targets.

A series of 2 (Labeling: L- vs. L+) X 2 (Target Type: Cheater vs. Non-Cheater) mixed factorials were performed on each of the previously described PD decision rates. Regarding subjects' cooperation rates, a significant Labeling X Target Type interaction was observed, $F\ (1, 55) = 6.566$, $p < .05$. Planned comparisons revealed that subjects were significantly more likely to cooperate against cheaters than non-cheaters in the L- condition, $t\ (28) = -2.186$, $p < .05$. However, while subjects were about 7.6% more likely to cooperate against non-cheaters than cheaters in the L+ condition, this difference was not significant. Still, the significant interaction demonstrates that subjects tended to cooperate more with non-cheaters when they had been given explicit information (via labeling) regarding who was and who was not likely to cheat.

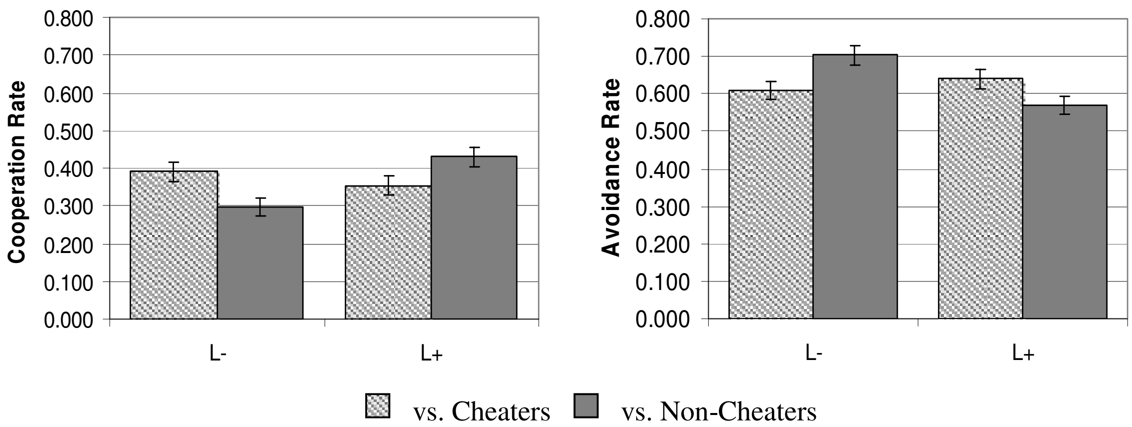


Fig. 1. Cooperation rates (left panel) and avoidance rates (right panel), divided by Stimulus Target Type and Labeling Condition ("L-" = without labels, "L+" = with labels).

Regarding subjects' defection rates, a significant Labeling X Target Type interaction was again observed, F (1, 55) = 6.089, $p < .05$. Planned comparisons revealed that subjects were significantly more likely to defect against non-cheaters than cheaters in the L- condition, t (28) = 2.675, $p < .05$. However, while subjects were about 5.7% more likely to defect against cheaters than non-cheaters in the L+ condition, this difference was not significant. Subjects' avoidance rates held a similar pattern, as the Labeling X Target Type interaction was again significant, F (1, 55) = 6.236, $p < .05$. Planned comparisons revealed that subjects were significantly more likely to avoid non-cheaters than cheaters in the L- condition, t (28) = 2.190, $p < .05$. While subjects in the L+ condition were about 7.0% more likely to avoid cheaters than non-cheaters, this difference was not significant. Still, the significant interactions revealed by factorials on both defection and avoidance rates demonstrates that, as hypothesized, subjects tended to be much better at avoiding cooperative gestures toward cheaters when they had been given explicit information (via labeling) regarding who was and who was not likely to cheat.

Cheater Avoidance: Explicit Knowledge is Required

Thus, the evidence presented here suggests that subjects were marginally capable of preferentially avoiding cheaters only when those cheaters had been labeled as such during exposure lists (i.e., upon the subjects' first perceptions of the cheaters' faces). There is no evidence to suggest that subjects were capable of avoiding cheaters when they were not privy to this explicit knowledge. In fact, if anything, subjects were particularly bad at "reading" faces for physiognomic signs of defection or cooperation, as they tended to cooperate with cheaters and to defect against cooperators in the L- condition; this pattern resulted in the subjects' exploitation by the cheaters as well as the subjects' failure to establish fruitful cooperative alliances with the non-cheaters. The notion that such explicit knowledge would be required for successful cheater avoidance should not be considered surprising in light of our intense interest in gaining such knowledge about other people, especially those of our own age, with whom we are particularly likely to interact (McAndrew & Milenkovic, 2002). We also take deliberate measures to protect our own reputations; Pinker (1997) has dubbed reputation as humankind's "most valuable possession" (p. 405) due to its pervasive influence on others' behavior toward us.

This pattern of evidence also tentatively suggests that modern humans more typically employ a strategy of cooperation that is more sophisticated and not as "nice" as tit-for-tat. That is, contrary to tit-for-tat, it does not seem that we must be exploited by a cheater in order to label that individual as a cheater; rather, we can glean this information from other sources, such as by observing the cheater or by gaining reputation information about the cheater. Armed with this knowledge, we can rely on it to avoid being exploited by that cheater. However, our strategy of cooperation does not appear to be so sophisticated that we can, in the absence of explicit knowledge about a potential cooperation partner, utilize physiognomic cues to direct our decisions regarding whether or not to cooperate.

Future Directions

In the present study, only still photographs were used as stimulus targets. Future research should experiment with the use of dynamic faces and audible voices to determine whether the cues inherent in these media offer any additional decision-making help to subjects. In addition, the methodology used to collect stimulus targets allowed subjects to know that they were being photographed; thus, they may have taken measures to "cover up" any facial signs of their previous defection, which would have made it more difficult for subjects playing simulated PD games to distinguish between cheaters and non-cheaters. The use of candid photographs could help overcome such an obstacle.

Future research should also continue to experiment with different methods of "cheater" classification. In the current study, cheaters were classified as such on the basis of only one PD decision. The use of iterated PD games or, alternatively, the Mach V Scale (Christie, 1970), could be useful in identifying more habitual "cheaters."

ACKNOWLEDGMENTS

This study was sponsored by a grant from Central Michigan University's Graduate Student Research and Creative Endeavors program. The authors also thank Dr. Toshio Yamagishi, Dr. Hajime Otani, and Dan Bracken for their assistance with this research.

REFERENCES

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books, Inc.

Axelrod, R., & Hamilton, W.D. (1981). The evolution of cooperation. *Science*, *211*, 1390-1396.

Christie, R. (1970). Scale construction. In R. Christie & F.L. Geis (Eds.), *Studies in Machiavellianism* (pp. 10-34). New York: Academic Press.

Dawkins, R. (1976). *The selfish gene*. New York: Oxford University Press.

Ekman, P. (1996). Why don't we catch liars? *Social Research*, *63*, 801-817.

Hamilton, W.D. (1964/1980). The genetical evolution of social behaviour. In J.H. Hunt (Ed.), *Selected readings in sociobiology* (pp. 7-30). St. Louis: McGraw-Hill.

Maynard Smith, J. (1982). *Evolution and the theory of games*. New York: Cambridge University Press.

McAndrew, F.T., & Milenkovic, M.A. (2002). Of tabloids and family secrets: The evolutionary psychology of gossip. *Journal of Applied Social Psychology*, *32*, 1064-1082.

Mealey, L., Daood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, *17*, 119-128.

Oda, R. (1997). Biased face recognition in the prisoner's dilemma game. *Evolution and Human Behavior*, *18*, 309-315.

Pinker, S. (1997). *How the mind works*. New York: W.W. Norton and Co.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior*, *24*, 290-301.

Yarmey, A.D., Yarmey, A.L., & Yarmey, M.J. (1994). Face and voice identifications in showups and lineups. *Applied Cognitive Psychology*, *8*, 453-464.

ODOR SENSITIVITY AS A FUNCTION OF GENDER AND SEXUAL ORIENTATION

Pamela Browder
Department of Psychology, Central Michigan University, Mt. Pleasant Michigan 48859 USA, browd1pe@cmich.edu

Abstract

Olfactory sensitivities were compared and contrasted in four groups consisting of gays, lesbians, heterosexual males and heterosexual females using a forced choice discrimination task. Androstenone, androstenol, and amyl acetate were chosen because males and females differ in their sensitivity to these odors. Gay males detected all odors better than heterosexual males. Lesbians were unable to discriminate as well as heterosexual females alone and when heterosexual females were combined with bisexual females. The most interesting finding was that the pattern of sensitivity showed more similarities between heterosexual females and gay males and between heterosexual females and gay males. Results suggest sexual orientation should be considered in studies looking for sex differences in olfactory sensitivities.

Sex differences in olfactory abilities are well documented, with women typically outperforming men in both identification and detection. (Gilbert & Wysocki, 1987, Koegla & Koster, 1974) One of the largest more in depth studies was Koegla and Koster (1974) documenting differences in favor of females in twelve out of the fourteen odors tested, no sex differences in favor of males were found. As with most studies the most pronounced differences occurred in the musk type odors several of which are now often referred to as putative human pheromones. The national geographic smell survey of over a million participants showed females were better in both detection and identification (Gilbert & Wysocki). Females outperforming males has been reported to occur in childhood (Beauchamp & Wysocki 1990), during puberty (Koegla & Koster, 1974), and into old age (Ship et al.). This superiority in olfactory tasks has been explained by some as an advantage in verbal abilities rather than increased perceptuary abilities, and this may be the case in instances of identification abilities; however, all of these studies have reported female superiority in threshold and discrimination tasks that are likely not dependant on verbal substrates.

In addition to the male/female differences other aspects of human sexuality like puberty, pregnancy, and ovulation have also been shown to have effects on olfactory sensitivity. Concerning puberty, Beauchamp and Wysocki (1990) found that anosmia or decreased sensitivity to androstenone increases with age with the largest effects occurring between the ages of 9-14 and 15-20, much more pronounced in males. The implications of these studies are that threshold levels for specific odors, putative human pheromones, change through puberty especially in males. Interestingly, sensitivity to the odor amyl acetate, not considered biological in origin, appears to be stable before, during, and after puberty (Beauchamp & Wysocki 1990; Koegla & Koster, 1974). Pregnancy was shown to decrease olfactory perception in the The National Geographic smell survey (Gilbert & Wysocki, 1987). Kathleen Dorries (1992) cites several studies in *Science of Olfaction* that show increases in olfactory sensitivity during ovulation or both ovulation and menses.

Recently, research has shown that sexual orientation can affect certain aspects of olfaction. Savik et al (2005) found variations in brain activation among gay males and heterosexual males, more specifically, finding that brain activation in gay males is more in line with heterosexual females than heterosexual males. Martine et al (2005) reported that sexual orientation effected ratings of body odor among gay, lesbian and heterosexual subjects. Given this research along with studies showing sexuality playing a role in human olfactory discrimination a study comparing and contrasting sensitivity levels among gays, lesbians, heterosexual males, and heterosexual females was conducted to look for additional sex related differences in olfaction. Two of the odors chosen have shown large sex differences in male/female studies and are putative human pheromones androstenone (urinous, musky), androstenol (milder musky). Amyl acetate (banana, fruity) was chosen because sex differences to this odor occur in childhood and sensitivity to this odor is not effected by exposure or puberty (Gilbert & Wysocki, 1987, Koegla & Koster, 1974).

Method

Subjects

Subjects were recruited from psychology courses and organizations with members that identify themselves as gay or lesbian. The participants were asked to establish their own sexual orientation by answering whether they were homosexual, bisexual or heterosexual. Additional questions were asked about current nasal congestion and use of allergy medication. All information was gathered on the back of their data sheet after they had completed testing. We only analyzed data for participants who reported no current nasal congestion and did not take daily allergy medication, this excluded 7 females and 7 males. In the female group N=27, 8 identified themselves as homosexual, 4 as bisexual and 15 as heterosexual, age ranged from 18 to 23 (M=19.8, SD=1.6). In the male group, 11 identified themselves as homosexual, 1 as bisexual and 8 as heterosexual. Because only one male identified himself as bisexual data was only analyzed for those who identified themselves as gay or heterosexual in the male group, N=19, age ranged from 18 to 23 (M=20, SD=1.4). All participants either received credit in an introductory psychology course or were paid five dollars.

Materials

Three odors were used: 16,(5α)-ANDROSTEN-3-ONE (Steraloids), 16,(5α)-ANDROSTEN-3α-OL (Steraloids), and Amyl Acetate (IFF). Each odor had six dilutions, for a total of eighteen presentations. All were diluted in food grade pure white mineral oil and presented in 6oz glass jars. Optimal dilutions were established in pre-trials. The initial dilution for 16,(5α)-ANDROSTEN-3-ONE was 15mg of white crystalline powder diluted in 10mL of mineral oil. A series of solutions were made using 2mL of the previous dilution mixed with 10 mL of white mineral oil. A total of 16 dilutions were made, from these; series 6, 8, 10, 12, 14, & 16 were used. The same procedure was used for amyl acetate, only the initial solution was .01mL diluted in 10mL of mineral oil and the series was made using 2.5mL of the previous dilution for a total of 18 dilutions, from these series 8, 10, 12, 14, 16, & 18 were used. The initial dilution for 16,(5α)-ANDROSTEN-3α-OL was 15mg of white crystalline powder diluted in 10mL of mineral oil. Only 14 dilutions were needed in this series using 4mL of the previous solution mixing it with 10mL of mineral oil, from these; series 4, 6, 8, 10, 12, & 14 were used.

Procedure

Testing took place in the same room over a six-day period. A forced choice method was used where target odors were placed on a paddle with three control odors (the control odor was also the solvent). There were a total of 18 trials, 6 dilutions of the 3 odors. Order in which the odors were presented was randomized for each subject and placement of the odorous solution was

randomized for each trial. Subjects were told to sniff each jar and indicate which one smelled different. They were instructed to avoid picking up the jars and to only smell each one jar once. If they were unsure of a choice they were told to guess. They also indicated the intensity and their confidence in correct choice for each trial. These were based on a 5 point scale with verbal descriptors. The scale was placed on the table of the testing area for subjects to refer to.
Testers marked the data sheets where the subjects indicated the odor was: either slot A B C D. Then indicated the ratings next to each trial. After each trial a cardboard partition was placed between tester and subject so testers could change odor and placement. Subjects received one point for every correctly identified odor. After testing was complete subjects were handed their data sheet and directed to a discrete location to answer the questions on back. This was done to insure anonymity and to increase honesty with respect to personal questions. In addition to gender and sexual orientation questions were asked about smoking, age, time of menses, birth control use, amount of cologne use, nasal congestion and daily use of allergy medication.

Results and Discussion

	Heterosexual Female *n*=15		Lesbian *n*=8		Heterosexual Male *n*=8		Gay Male *n*=11	
	M	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
Androstenone	4.47	(1.3)	3.25	(1.2)	4.38	(1 .3)	4.64	(1.3)
Androstenol	4.93	(1.3)	3.75	(1.2)	4.13	(1.4)	4.46	(1.4)
Amyl Acetate	4.93	(1.3)	4	(1.3)	4.4	(1.4)	5.18	(1.2)
All Odors	14.3	(2.6)	11	(3.2)	12.9	(3.3)	14.3	(3.4)

Table 1.

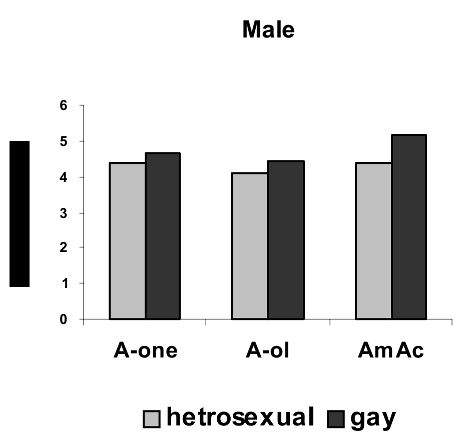


Figure 2. ***Androstenone*** *(A-one)*
Androstenol (A-ol)
Amyl Acetate (AmAc)

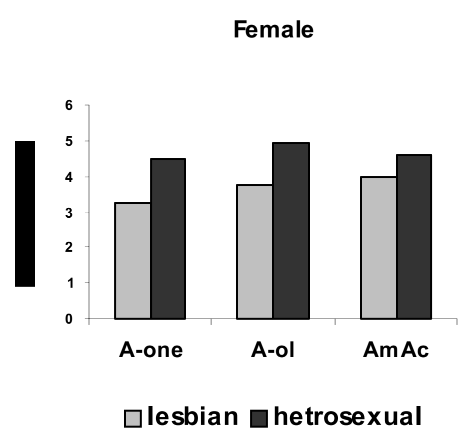


Figure 2. ***Androstenone*** *(A-one)*
Androstenol (A-ol)
Amyl Acetate (AmAc)

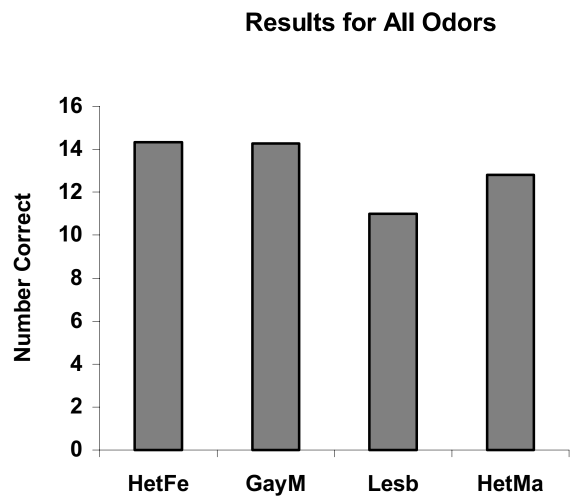


Figure 4. ***Number Correct out of 18.***
Heterosexual Female (HetFe), Gay Male (GayM)
Lesbian (Lesb), and Heterosexual Male (HetMa)

A review of studies looking at sex differences in olfaction will reveal varied results that are highly dependant on the odor being tested. For example, studies have shown no sex differences for nbutanol (Oberg et al. 2001), milder differences for amyl acetate and the most pronounced in musk odors and putative human pheromones like androstenone (Koegla & Koster, 1974). In this study, ANOVA revealed lesbians had significantly lower discrimination abilities than heterosexual females ($F_{1,21} = 7.2$ $p < .05$) and when heterosexual females were combined with bisexual females ($F_{1,25} = 6.7$ $p < .05$). When the odors were analyzed individually, significant differences between these groups were found in androstenone and androstenol ($p < .05$), but not in amyl acetate $p = .16$. It is possible that amyl acetate was not significant due to the fact that there were a relatively small number of subjects, because the differences for amyl acetate in larger male/female studies is not as pronounced as those in putative human pheromones and would be less likely to detect with fewer subjects (Koegla & Koster, 1974). Or it could be that sexual orientation only plays a role in discrimination for these odors. Savic et al (2005) found a difference between gay and heterosexual males in olfactory substrates for the processing of putative human pheromones as opposed to odors that are not biological in origin.

The overall pattern of sensitivity shows that gay males are performing similar to heterosexual females and lesbians are more inline with heterosexual males (Fig. 4). The sensitivities between gay males and heterosexual males did not differ enough to reach statistically significant. However, gay males were more sensitive to all three odors (Fig. 1). A weakness of this study was that subjects had to identify either as heterosexual, bisexual or gay. It is possible that a the Kinsey Scale, with a continuum of sexuality, may have provided a more fine gradient among males as only one identified himself as bisexual. In this study women were more likely to identify as bisexual than men. Additionally, dilutions at higher thresholds may have provided more data as some subjects correctly discriminated all odors and/or all in an individual odor. Although, it should be noted, that while not the focus of the study, Savic et al. (2005), found no threshold differences between heterosexual women, heterosexual men or gay men (12 in each group) for 4,16-androstadien-3-0l and estra-1,3,5(10), 16-tetraen-3-0l. Lesbians were not included in this study

Determining the reasons heterosexual females performed significantly better than lesbians will be complex and requires additional research. Several possibilities exist. Laboratory studies show that systematic exposure to androstenone can increase sensitivity levels (Beauchamp & Wysocki 1990). It is conceivable that heterosexual females may be exposed to more androstenone. Androstenone is found in human auxiliary secretions in much higher levels in males. (cited in Dorries 1992) However, exposure to males was not assessed and differences in factors like housing may have been similar among subjects as all were young college students. Studies correlating increased sensitivity to androstenone in females in relation to contact with males were not found. In addition, at least some androstenone sensitivity is genetic; twin studies have shown that sensitivity to androstenone was concordant for 100% of identical twins and 61% of fraternal twin (Beauchamp & Wysocki, 1990).

Recent studies showing sexual orientation playing a role in olfaction warrants additional research. Utilizing preexisting knowledge about sex differences in specific odors may offer insight into the specific function sexual orientation plays in olfaction. For example, if differences are found in odors that have not shown general male/female differences it may be due to general abilities or deficits in olfactory perception. However, finding sexual orientation differences among odors that are typically sex differentiated could imply that hormones are involved because these are cited as the primary cause for differences related to sexuality (Dorries 1992). Additional research into the specific role sexual orientation plays in olfaction will provide valuable information not only into the role hormones play in odor sensitivity but possibly into the behavioral aspects of olfaction as well.

References

Beauchamp G. & Wysocki C (1990). *Perception of the Odor of Androstenone: Influence of Genes, Development, and Exposure.* Taste, Experience and Feeding ed. Capaldi & Powley. APA, Washington

Dorries, K.M. (1992). Sex differences in olfaction in mammals. In M.J. Serby & K.L. Chobor (Eds.) *Science of Olfaction*. New York: Springer Verlag.

Gilbert, A.N., & Wysocki, C. J. (1987). The Smell Survey Results. *National Geographic, October* 514-525.

Koegela, H.S., & Koster, E. P. (1974). Some experiments on sex differences in odor perception. *Annals of the New York Academy of Sciences, 237*, 234-246.

Martins, Y., Preti, G., Crabtree C., Runyan, T., Vainius, A., & Wysocki C (2005) Preference for Human Body Odors Is Influenced by Gender and Sexual Orientation. *Psychological Science* 694-701.

Savic, I., Berglund, H., & Lindstom P. (2005). Brain Response to Putative Human Pheromones in Homosexual Men. *Proceedings of the National Academy of Science.*

Oberg, C., Larsson M., Backman, L. (2001). Differential Sex Effects in Olfacltory Functioning: The Role of Verbal Processing. *Journal of the International Neuropsychological Society.* 691-698.

Ship, Johnathan A., Pearson, Jay D., Cruise, L. J., Brant, L. J. & Metter J. E. (1996). Longitudal changes in smell identification. *Journal of Gerontology: MEDICAL SCIENCE Vol. 51A, (2),* M86-M91.

FUNCTIONAL EQUIVALENCE AND TRADE-OFF BETWEEN BRIGHTNESS AND SATURATION IN YOUNG-ADULTS AND OLD-AGED

Nuno Colaço[1], António M . Diniz[2], A. M. Oliveira[3]
[1]ULHT, Lisboa; [2]ISPA, Lisboa; [3]FPCEUC, Coimbra
nuno.colaço@ulusofona.pt

Abstract

Colour has been physically depicted as the result of three underlying dimensions: hue, saturation and brightness. The issue of their perceptive separability is seldom addressed in explicit terms. However, the fact that they can be physically distinguished is mute regarding the possibility of their subjective discriminability. This study explores the functional equivalence of two of these dimensions, brightness and saturation, by means of a SDT approach, both in young-adults and elderly people. Outcomes illustrate extended trade-off between dimensions, with changes in brightness being fully matched by changes in saturation, and vice-versa.

Spectral colour is physically composed of three basic dimensions, grossly definable as follows. Hue: the quality that supports colour's distinctiveness. Brightness: the quantity of light. Saturation: the amount of grey that renders a colour pale or vivid. The issue of their perceptive separability is not often address in explicit terms, although it is a basic acknowledgment that different physical configurations of the spectral space, embodying distinct physical properties, can be perceived as corresponding to a single, invariant colour (Sekuler & Blake, 1990). At the level of subjective processing where visual perception stands, colour is treated in a fairly independent way from its physical composition (Boker, 1997).

As far as brightness and saturation are concerned, clear suggestions can be found that they are not subjectively envisaged as distinct dimensions. In a study of the "harmony of colour" through conjoint measurement, Pietrs (1979) reported that changes in saturation were subjectively perceived as changes in hue or in brightness. From a neurophysiological standing, the opposition of parvocelular and magnocellular pathways appears to subserve the one between automatic detection of brightness changes and attentive detection of colour changes (meaning, change in hue), while no equivalent specific anatomical pathway is known regarding saturation (Theeuwes, 1995). Also, known functional asymmetries among colour dimensions concern the dimensions of hue and brightness – varying brightness interferes with hue discrimination, while the converse is not true -, with no functional separate room for saturation (Callaghan, 1990).

The present study originated as a pilot check on the need to give separate consideration to brightness and saturation (as design factors) in an experiment concerned with the visual search for colours in young and old-aged subjects. Despite little doubts that both dimensions would be subjectively handled as one, a preliminary experiment was thus conducted, involving small samples of young adults and elderly, with the aim of assessing the extent of functional equivalence (rather than of reported phenomenological overlap) between those two dimensions.

The approach taken followed a two-steps procedure, conjoining classical psychophysical threshold determination (method of limits) and Signal Detection Theory (same-different paradigm). The first step consisted in determining points of subjective equality (PSE: hereafter also used to refer to Stimuli of Subjective Equality) to blue and yellow standard stimuli (SS), obtained in one case by varying saturation values, and on the other by varying brightness values. These PSE stimuli were then used, at a second step, to build pairs of subjectively equal stimuli (PSE_{bright}/SS; PSE_{sat}/SS), contrasted with pairs of subjectively different stimuli (selected from those immediately below or above the threshold differences) (SS/DS; PSE_{bright}/DS; PSE_{sat}/DS). Final outcomes were dictated by the statistical analysis of signal detection indices obtained from same-different judgments, issued by subjects while facing these pairs. Predictions could be simply equated: no emerging differences between the different sorts of "same" pairs, as well as between the several sorts of "different" pairs (even if some degree of qualification may be needed in this last case), would signal extended functional equivalence (and also extended subjective trade-off) between saturation and brightness.

Method

Subjects. Four young adults (mean age = 23), and four old-aged subjects (mean age = 81), all with normal or corrected to normal vision, participated in the study. Records of cognitive deficit, assessed in the elderly group by the Portuguese version of the MMSE scale (Guerreiro, Silva & Botelho, 1994), or problems in colour vision assessed in both groups by the *Ishiara's Test for Colour Deficiency* (Ishiara, 2003) were used as exclusion criteria.

Stimuli. **I.** Yellow and blue standard stimuli (SS) were chosen and defined in RGB coordinates. Brightness values for the standards corresponded to 100% (brightness variation can go above that), while saturation values were fixed at 70%, to allow for both upward and downward variation of the dimension. Two sets of 13 stimuli varying respectively in brightness and saturation by steps of 5% were built for each colour (yellow and blue), to be used on the determination of PSE and threshold differences. **II.** 120 pairs of stimuli were built for each colour, involving SS, PSE, and close-to-liminar above-threshold stimuli (DS), to be used in a same-different discriminating task. With the exception of SS, all the other stimuli were individually tailored for each subject

Design and procedure. **I.** PSE and DS determination proceeded through 10 alternately ascending and descending presentations of the variable stimuli, accompanied by the simultaneous presentation of the SS (method of limits); subjects judged whether the variable stimulus was higher, lower or equal regarding the standard, for either brightness or saturation. Half the subjects started by the yellow colour, the remaining half by the blue colour. For those starting with a specific colour, half started by the "saturation block", the others by the "brightness block". PSE was determined for each subject, for brightness as well as for saturation, as the mean of all four difference threshold points; upper and lower difference thresholds were calculated as the mean of the two upper and of the two lower threshold points, respectively. The DS (different stimuli) were taken as the ones standing immediately above the upper difference thresholds or below the lower difference thresholds.
II. Same-different task: 60 "same"pairs (20 SS/SS; 20 PSE_{bright}/SS; 20 PSE_{sat}/SS) and 60 "different" pairs (20 SS/DS; 20 PSE_{bright}/DS; 20 PSE_{sat}/DS) for each of the two colours (blocked) were randomly presented. The ordering of colour blocks was again evenly balanced across subjects. Subjects sat at 53 cm from a computer display in a closed room lightened by a Colour Daylight 100 watts lamp; they were asked to issue a same-difference judgment by correspondingly pressing the left or right mouse button. Data analysis concerned "proportion of correct answers" for each kind of "same" and "different" pairs, d'_{sd} (sensitivity parameter for same-different designs: Macmillan & Crelman, 1991) and "maximum proportion correct" (Macmillan & Crelman, 1991), again for each sort of stimuli pairs.

Results

Tables 1 and 2 summarize the binomial comparisons for all sorts of "same" and "different" pairs, both for young adults (table 1) and old-aged subjects (table 2), and for both hues (yellow and blue).
The criteria used for comparison were SS/SS results when "same" pairs were involved (PSE_{bright}/SS; PSE_{sat}/SS) and SS/DS results when "different" pairs were being considered (PSE_{bright}/DS; PSE_{sat}/DS). This allowed testing for the functional equivalence of SS and the PSEs obtained from brightness and from saturation (both upper and lower DS were evenly used in "different" pairs; this compensates in principle for the downward and upward shifts of the PSEs regarding the SS, which might otherwise endanger the comparisons).

Results concerned, in one case [(a)], observed correct proportions; in another case [(b)], "maximum proportion correct" - i.e., the correct proportion that a subject with a given *d'* (calculated from the proportion of hits and false alarms) would obtain if he used a neutral, "non biased" response criterion. This last index provides a way of rendering subjects' performances homogeneous as to factors extraneous to discriminative sensitivity.

As can be seen, if some deviations from the null hypothesis prediction were observed with "proportion correct" (namely in the elderly group, for the pair PSE_{bright}/DS), they almost completely vanish when "correction" for criterion "bias" was introduced (exception made for one young subject on the PSE_{bright}/DS pair). Most of the deviations concerned the "different" pairs, more vulnerable to the possible instability of DSs and the least fit to a "null hypothesis" approach. This pattern of results remains essentially the same in both groups and for either colour (yellow and blue).

Finally, concerning d'_{sd}, the same findings were also replicated. With one exception ("different" pairs, elderly group, yellow hue: sig. = .037), no other significant differences occurred between types of "same" pairs or between categories of "different" pairs for either hue (yellow, blue) or group (asymp. sig for Friedman Test > .108); comparisons between averaged *d'* for "same" and for "different" pairs always turned up to be at least marginally significant (Wilcoxon signed ranking test: sig. = .068). "same" pairs never distinguished significantly from 0 (zero) (sig > .223).

Table 1. Binomial tests: young-adults

	Yellow				Blue			
	Binomial (a)		Binomial (b)		Binomial (a)		Binomial (b)	
	p>.05	$p\leq$.05	p>.05	$P\leq$.05	p>.05	$p\leq$.05	p>.05	$p\leq$.05
Comparison criterion - SS/SS								
PSE_{bright}/SS	4	0	4	0	4	0	4	0
PSE_{sat} /SS	4	0	4	0	4	0	4	0
Comparison criterion - SS/DS								
PSE_{bright}/DS	4	0	4	0	3	1	3	1
PSE_{sat} /DS	3	1	4	0	4	0	4	0

Table 2. Binomial tests: elderly

	Yellow				Blue			
	Binomial (a)		Binomial (b)		Binomial (a)		Binomial (b)	
	p>.05	$p\leq$.05	p>.05	$p\leq$.05	p>.05	$p\leq$.05	p>.05	$p\leq$.05
Comparison criterion - SS/SS								
PSE_{bright}/SS	4	0	4	0	3	1	4	0
PSE_{sat} /SS	3	1	4	0	4	0	4	0
Comparison criterion - SS/DS								
PSE_{bright}/DS	3	1	4	0	2	2	4	0
PSE_{sat} /DS	4	0	4	0	3	1	4	0

(a) – number of subjects showing *sign.* and *n.s.* differences (0.05) for "proportion correct";
(b) – number of subjects showing *sign.* and *n.s.* differences (0.05) for "maximum proportion correct"

Discussion

The reported findings lend strong support to the view that the two physically distinguishable dimensions of saturation and brightness are actually functionally and subjectively handled as one. This general conclusion is in accordance with number of suggestions found in literature (e.g. Pietrs, 1979).

Several methodological features of the design help ensuring the soundness of results. Among them is the tailored nature of the stimuli. Every PSE and DS employed was specific to each individual (also, data analysis at the individual subject level was systematically pursued). A second one is the use of signal detection sensitivity parameters (d' and "max. proportion correct") as a basis for comparisons: this was instrumental in preventing the confounding with criterion parameters that affects binomial comparisons of "percentage correct". Thirdly, the replication of the same results on two different samples – young-adults and elderly –, and for two colours with known distinct profiling regarding the effects of age (yellow and blue) argues for the robustness of the findings.

However, some drawbacks should also be noticed. The rationale of the approach makes conclusions dependent on establishing the "null hypothesis". Also, the number of subjects sampled is small (a consequence of the pilot scope intended for the study). Desirably, than, the same conclusion should be reached through more constraining approaches to the perceptual separability of these dimensions (such as interference studies, or their conjoint factorial manipulation), relying moreover on larger samples. A crucial limit to the feasibility of such approaches may nonetheless exist if, at both the functional and phenomenological levels, brightness and saturation totally and fully overlap.

References

Boker, S. (1997). A measurement of the adaptation of color vision to the spectral environment. *Psychological Science, 8* (2), 130-134.

Callaghan, T. (1990). Interference and dominance in texture segregation. In D. Borgan (Ed.), *Visual search* (pp. 81-87). New York: Taylor & Francis.

Guerreiro, M., Silva, A. & Botelho, M. (1994). Adaptação à população portuguesa da tradução do Minimental State Examination (MMSE). *Revista Portuguesa de Neurologia, 1*, 9-10.

Ishihara, S. (2003). *The series of plates design as a test for colour-deficiency.* Tokyo: Kanehara Trading Inc.

MacMillan, N., & Creelman, C. (1991). *Detection theory: A user's guide.* Cambridge, MA: Cambridge University Press.

Pietrs, J. (1979). A conjoint measurement approach to color harmony.*Perception & Psychophysics, 26*, 4*,* 281-286.

Sekuler, R., & Blake R. (1990). *Perception.* New York: McGraw-Hill.

Theeuwes, J. (1995). Abrupt luminance change pops out; abrupt color change does not. *Perception & Psychophysics, 57*(5)*,* 637-644.

CORRELATION BETWEEN THE VISUAL AND TACTILE PERCEPTIONS: A STUDY WITH BLIND PEOPLE THROUGH STUFFED ANIMALS

Custódio, Vagner Sérgio and Ferraz, Marcelo Antonio
São Paulo State University at Rosana and University of São Paulo at Ribeirão Preto, Brazil
vagner@rosana.unesp.br; amefe@hotmail.com.br

Abstract

Stuffed animals and their respective ratio, forms and textures, have been judged through the tactile perception for blind people using scales of category and ranking with relation to the beauty, danger and repulse, the results being compared with a control group with vision that executed the task through pictures. Results indicated positive correlation between the visual and tactile judgments.

According to (Ramos-Estebanez; Machii; Merabet; Romei; Fregn; Rizzo; Pascual-Leone, 2004), the brain is always using all the perceptions to create a mental scene of the situation. In this perspective, this work has as a goal to prove that the Sensorial Safari, activity spread through the Safari Club International, is a viable activity for blind people, having as hypothesis that the tactile perception of stuffed animals possesses a satisfactory accuracy, since according to Millar (1994) the blind person possesses space information proceeding from different sensorial modalities during the exploration of the space, that are memorized and organized in mental representations. (Kerr 1983; Landau; Spelke; Gleitman, 1984; Leonard; Newman 1987; Millar, 1999).

This applied research, is justified due to difficulty of the appreciation of wild beasts by blind people, since many of these animals are dangerous, what disables the tactile perception in natural environments in a way that would guarantee the security conditions. Put another way, the simple verbal exposition on sizes, ratio, forms, and texture of wild beasts, cannot correspond adequately to the real image, mainly in congenital blind people who never had the visual perception. Therefore this work aims at to prove trustworthily, through psychophysics methods as the scale of category and ranking that stuffed animals can be a viable form of assisting in the space representation of these animals in a more trustworthy and hedonic way, promoting activities of ecotourism and environmental education with this population.

Method

Participants: Control Group (CG) - Twelve people (5M and 7F), inhabitants of Presidente Prudente, Brazil, participated as voluntary in this study. Their ages varied between 17 and 64 years (average=30.6 years of age). All declared to possess good vision with or without correction, and they did not know the intentions of the experiment.

Experimental group (EG) - Twelve people (9M and 1F), carriers of visual special necessities (Total Blindness), attended in the Philanthropic Association for the Protection of Blind People of Presidente Prudente - Brazil, being 2 congenital and 10 adventitious ones. Their ages varied between 30 and 71 years (average=54.5 years of age). All participated voluntarily in this study and they also did not know the intentions of the experiment.

All the participants had signed a declaration of assent also written in Braille clarifying as to the participation in the experiments in agreement with the effective norms for execution of experiments with human beings in Brazil, being that the research was approved by the committee of ethics of São Paulo State University.

Material and equipment. 10 stuffed wild beasts yielded by the Environmental Police of the State of São Paulo/Brazil, chosen randomly in the following order: Monkey, anteater, wolf, armadillo, deer, alligator, wild pig, jaguar, anaconda, jabiru.

Laptop Computer Pentium 4; Laser printer HP Jet 4100 PCL6; Multi-functional Lexmark X1185; Word Software; Excel Software; SPSS (Statistical Package will be the Science) Software; Digital photographic machine Sony.

1 form containing spread sheets for execution of the experiment as well as specific instructions for the fulfilling of each psychophysics test.

Color Photographs of all the above-mentioned stuffed animals.

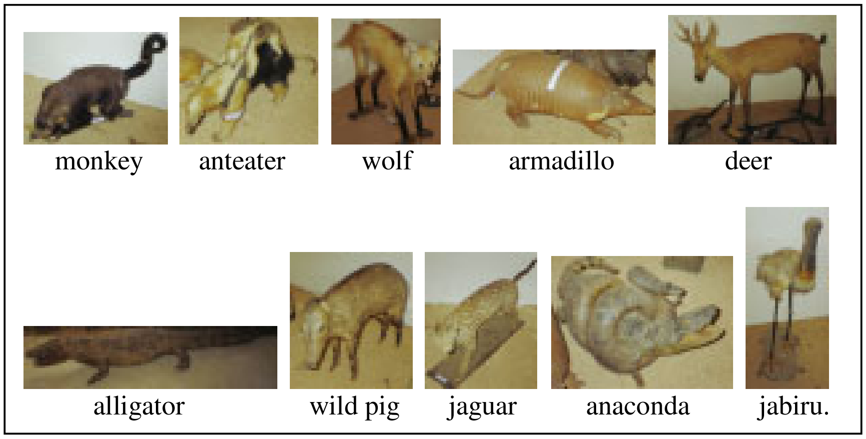


Figure 1. Photographs of all the stuffed animals used in the experiment.

Procedure. The methods of scale used were estimation of categories and ranking. On the first method, the task of the subject consisted of designating a number between 1 and 7 to each animal that was referring the beauty, danger and repulse. Module or stimulation standard had not been previously assigned, being that each subject established only 1 estimate, for each animal.

In the second method, the task of the subjects was to rank decreasingly all the animals presented in the experiment. Thus, if the subject judged that an animal possessed greater beauty among the others, it would have to designate the ranking 10. If the subject judged another animal has a lesser beauty, it would have to designate it as ranking 9 and so on until arriving at number 1.

This procedure was also used for the attributes danger and repulse. The application of the tests was carried through in the form of interview, and all the instructions given to the subjects, were written in the test form, and were read in an identical way to all the participants.

All the subjects made the judgments individually being that the control group carried through the judgment through photographs, in a room with a table for the application of the experiment. In the experimental group the experiment was applied in a room of Environmental Education of the São Paulo State University in Presidente Prudente which contained various stuffed animals. In the experimental group, as in the control group, the researcher restricted himself to the information contained in the instructions of the form, being that the participants had been able to appreciate the animals as long as they found necessary in order to carry through the judgments.

Results

In Table 1 the arithmetic means of the estimates of categories of the beauty, danger and repulse attributes are presented, of each stuffed animal judged by the 12 participants of the control group and the 12 participants of the experimental group.

Table 1. Arithmetic means of results of the tests estimates of categories in both groups with regards to beauty, danger and repulse.

animals	beauty control group	danger control group	repulse control group	beauty experimental group	danger experimental group	repulse experimental group
monkey	3,7	3,6	3,0	4,5	2,8	3,1
anteater	5,9	2,6	2,0	4,6	4,4	2,8
wolf	3,8	5,3	4,1	5,3	4,9	2,7
armadillo	4,5	2,1	3,1	5,3	2,0	2,8
deer	5,7	3,1	2,0	6,0	2,3	2,3
alligator	3,8	6,7	5,4	4,3	5,7	3,2
wild pig	3,0	5,0	3,9	4,8	5,4	2,8
jabiru	6,5	4,6	2,2	5,8	4,4	2,6
jaguar	4,9	6,9	3,9	5,0	5,7	3,9
anaconda	4,6	3,7	3,6	5,5	3,5	2,8
anaconda	4,6	3,7	3,6	5,5	3,5	2,8

It can be observed in this table, that in the method of estimates of categories, with regards to beauty jabiru (EC=6,5) was considered by the control group as the most beautiful animal, followed for anteater (EC=5,9). The experimental group, through tactile perception, judged deer (EC=6,0) judged as the most beautiful followed by jabiru (EC=5,8). With regards to danger, the control group judged through photographs the jaguar (EC=6,9) as being the most dangerous one, the alligator (EC=6,7) in second place. In the experimental group of blind people occurred a tie between alligator and jaguar (EC=5,7). With regards to repulses, the control group judged the alligator (EC=5,4) as being most disgusting, and the experimental group judged jaguar(EC=3,9) as being most repulsive.

These results demonstrate that in the attributes of beauty and danger, both groups presented similar judgments, and with regards to repulse, the groups judged different animals.

In Figure 2, the arithmetic means of the estimates of categories are projected comparing the judgments of both groups with regards to beauty, and the attribute danger is represented in Figure 3 and repulse in Figure 4.

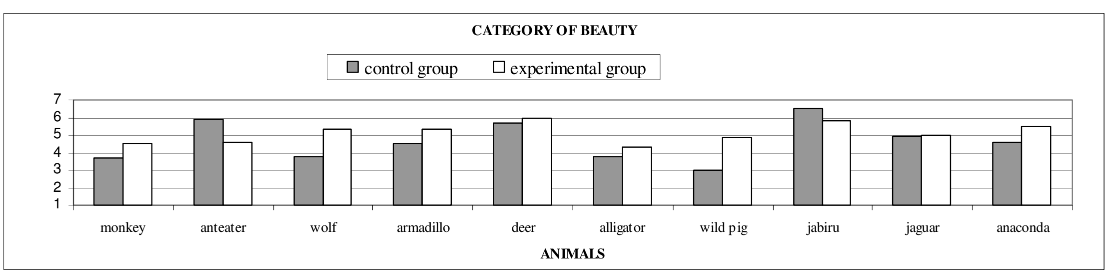


Figure 2. Arithmetic means of the estimates of categories comparing the attribute beauty between the control group and the experimental group.

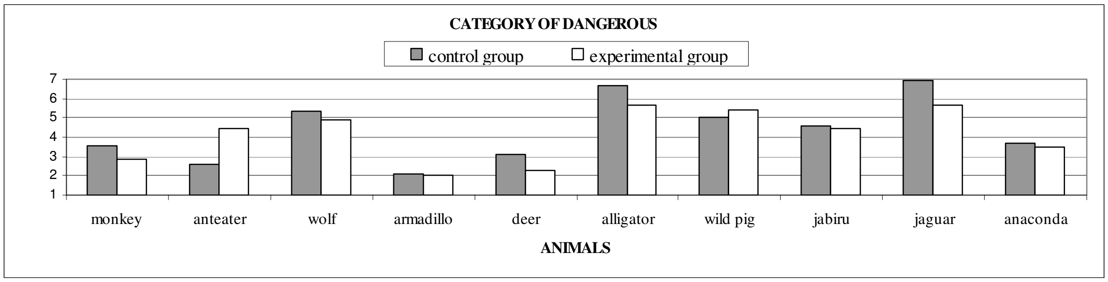


Figure 3. Arithmetic means of the estimates of categories comparing the attribute danger between the control group and the experimental group.

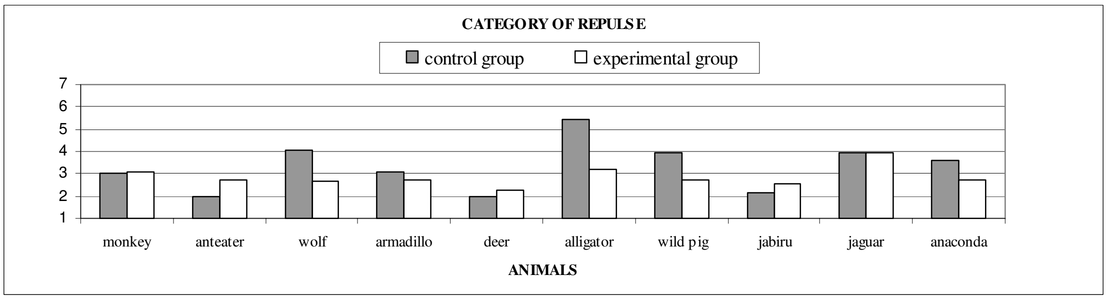


Figure 4. Arithmetic means of the estimates of categories comparing the attribute repulse between the control group and the experimental group.

Table 2 presents measured Arithmetic averages of ranking of both groups with relation to the beauty, danger and repulse atributes, being that the animal best judged was attributed a concept 10 and the worse judged concept 1.

Table 2. Arithmetic means of results of ranking in both the groups with regard to beauty, danger and repulse.

animals	beauty control group	danger control group	repulse control group	beauty experimental group	danger experimental group	repulse experimental group
monkey	4,8	4,5	5,5	4,8	4,3	6,9
anteater	7,8	3,3	3,0	5,1	6,3	5,1
wolf	4,9	7,2	7,0	5,6	7,0	5,3
armadillo	5,3	2,4	5,4	5,9	1,8	5,8
deer	6,6	3,7	3,4	7,3	4,1	2,8
alligator	3,9	8,6	8,1	4,8	6,2	5,7
wild pig	3,2	5,5	7,9	3,6	5,9	7,4
jabiru	6,9	2,5	2,6	8,2	1,9	2,8
jaguar	6,9	8,7	3,1	8,2	7,8	5,3
anaconda	2,3	8,7	8,9	2,3	9,3	7,8

With relation to ranking the control group classified anteater (RK=7,8) in first place and in the attribute beauty, different from the experimental group that classified jabiru and the jaguar with (RK=8,2). In the attribute danger, the control group classified the jaguar and anaconda (RK=8,7), and the experimental group only classified the jaguar (RK=7,8), and in repulse the animal chosen for both the groups was anaconda CG (RK=8,9) and EG (RK=7,8).

According to results, only in the attribute beauty, the groups had presented dissimilar judgments.
In Figure 5, the arithmetic means of ranking are projected comparing the judgments of both groups with regards to beauty, the attribute danger being represented in Figure 6 and repulse in Figure 7.

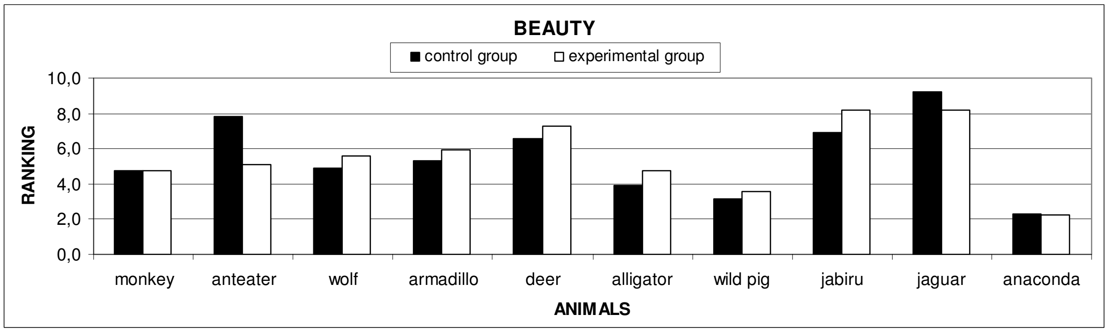


Figure 5. Arithmetic means of ranking comparing the attribute beauty between the control group and the experimental group.

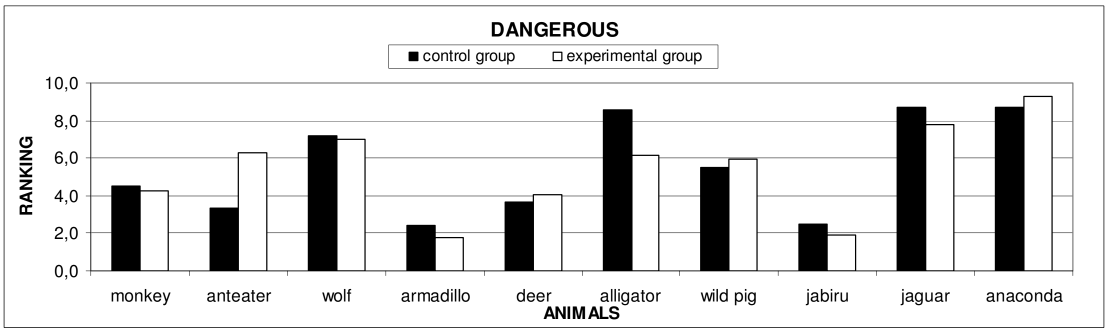


Figure 6. Arithmetic means of ranking comparing the attribute danger between the control group and the experimental group.

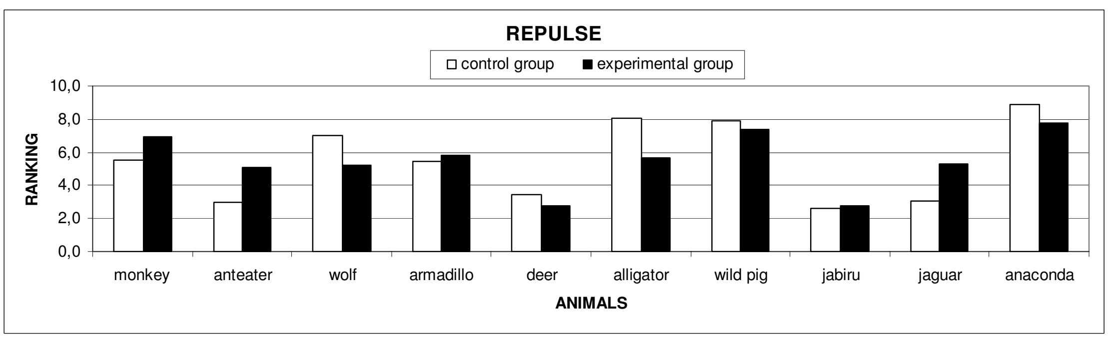


Figure 7. Arithmetic means of ranking comparing the attribute repulse between the control group and the experimental group.

The Table 3 and Table 4, show the results of the Spearman Rank Order Correlations, calculated between the ordinances of the estimates of category and ranking of both the groups with relation to the beauty attributes, danger and repulse. CABEG (Categories of Beauty Experimental Group), CABCG (Categories of Beauty Control Group), CADEG (Categories of Danger Experimental Group), CADCG(Categories of Danger Control Group), CAREG(Categories of Repulse Experimental Group), CARCG(Categories of Repulse Control Group). RKBEG (Ranking of Beauty Experimental Group), RKBCG (Ranking of Beauty Control Group), RKDEG (Ranking of Danger Experimental Group), RKDCG (Ranking of Danger Control Group), RKREG (Ranking of Repulse Experimental Group), RKRCG (Ranking of Repulse Control Group).

Table 3. Scores of Spearman Rank Order Correlations of estimative of categories

Spearman Rank Order Correlations (carkce.sta)

MD pairwise deleted

		Valid N	Spearman R	t(N-2)	p-level
CABEG &	CABCG	10	0,539634	1,812943	0,107406
CADEG &	CADCG	10	0,490721	1,592957	0,149835
CAREG &	CARCG	10	0,89026	5,528657	0,000555

Table 4. Scores of Spearman Rank Order Correlations of Ranking

Spearman Rank Order Correlations (newrak.sta)

MD pairwise deleted

		Valid N	Spearman R	T(N-2)	p-level
RKBEG &	RKBCG	10	0,86587	4,895461	0,001201
RKDEG &	RKDCG	10	0,851068	4,584623	0,001791
RKREG &	RKRCG	10	0,765961	3,3699	0,009787

Discussion

According to the results, ranking was relatively similar between the groups. It can be attributed to this, subjective cognitive criteria, of the representations that the subjects already possessed through cultural information that the animal represents, and consequently, the judgment possesses a lesser sensorial and percipient influence.

For the estimates of categories to which attributes of values would have to be related, it demonstrated a low correlation for the attributes beauty and danger, once that the percipient and sensorial factor can have influenced the judgments using different sensorial canals.

In this perspective, the tactile perception demonstrated other cognitive representations for the blind subjects that made them judge beauty and danger in different ways than they already possessed of cognitive knowledge on the animals. However in the attribute repulse this did not occur since this factor brings with it a bigger demand of representations that are acquired in cognitive way.

According to results, the hypothesis that the tactile perception of blind people in stuffed animals possesses a satisfactory accuracy for the development of activities of ecotourism and environmental education is confirmed, and this opens perspectives for the development of other experiments applied with this thematic.

Acknowledgements

This research was supported by FUNDUNESP – Foundation for the Development of UNESP

References

Kerr, N. H. *The role of vision in visual imagery experiments: evidence from the congenitally blind.* Journal of Experimental Psychology: General, v. 112, p. 265-270, 1983.

Landau, B., Spelke, E. & Gleitman, H. *Spatial Knowledge in a young blind child.* Cognition, v. 16, p. 225-60, 1984.

Leonard, J. A. & Newman, R. C. *Spatial orientation in the blind.* Nature, v. 215, p.141-314, 1967.

Millar, S. Veering re-visited: noise and posture cues in walking without sight. Perceptíon, v. 28, p. 765-780,1999.

Pashler, H. *Stevens' handbook of experimental psychology*. John Wiley & Sons, v. 4, n. 3, 2002.

Ramos-Estebanez C, Machii K, Merabet C, Romei V, Fregn F, Rizzo Jf, Pascual-Leone A. *Tactile modulation of occipital cortex function: a TMS study.* 5th International Multisensory Research Forum, Barcelona, 2004.

GENDER DIFFERENCES IN POWER FUNCTIONS (N-VALUES) FOR LOUDNESS

Lisa M. D'Alessandro,[1] Elad Sagi,[2] Kenneth H. Norwich[1]

[1] *Institute of Biomaterials & Biomedical Engineering, University of Toronto*
[2] *DeVault Otologic Research Laboratory, Indiana University School of Medicine*
l.dalessandro@utoronto.ca, esagi@iupui.edu, k.norwich@utoronto.ca

Abstract

Experiments were conducted to determine the ability of participants to identify intensities of tones to the nearest decibel. Errors in identifications were found to be normally distributed. The greater the range of intensities over which the tones could lie, the greater the variance of the distribution. Standard deviation increases linearly with range and it is shown that the slope of this line is a measure of the loudness exponent, *n*. Values of *n* for females are statistically greater than those of males. This result was anticipated from decreased loudness tolerance of females found in earlier studies.

Participants in this study were required to identify the intensity of tones at a given constant frequency, presented to them randomly through headphones. He or she was told only the range over which the random tones would be selected: 1 – 9, 1 – 27, …, 1 - 90 dB HL. The participants made errors in identification, and the errors were governed by the normal distribution. The variance of the normal curve of errors was a monotonic function of the range of tone intensities used in a given experiment. That is, if we tested over the range 1 – 9 dB HL, variance was s_1^2; if we tested over the range 1 – 27 dB HL, variance was $s_2^2 > s_1^2$ etc. We found that when standard deviation, s (in decibels), was plotted against range (also in decibels), a straight line resulted. We argue that the slope of that straight line bears a simple relationship to the power function exponent for a tone at the given frequency.

Gender differences in the auditory system have been studied by other researchers (McFadden, 1998, Don et al., 1993) and yet no difference in power function exponents for loudness was recorded. However, McGuinness (1974), in studying gender differences in loudness tolerance noted that females have a mean maximum comfort level 8 dB lower than that of males. She also inferred that, "for women the slope of the loudness function would be steeper than that of men." We were able to study this phenomenon using the intensity-identification process introduced above.

Experimental Methods

At the beginning of an experimental session, participants were informed of the range over which the experimental tones could lie. They were then trained to identify tones to the nearest decibel in two stages. First, participants would administer the tones to themselves from the stimulus range to be tested that session. Second, the experimenter would present a tone randomly selected from the stimulus range, and the listener would estimate the intensity to the nearest decibel. Afterwards, feedback would be provided. This process was continued until the listener felt comfortable making identifications. The first form of practice lasted about 5 minutes while the second took no more than 10 minutes. Feedback does not improve performance, as discovered by Mori and Ward (1995). Hence, during the experimental session no trial-by-trial feedback was offered. A typical experimental session lasted 30 minutes. Every 20 seconds, a tone was presented and the participant would provide his/her best estimate of the tone intensity to the nearest decibel. Taken together, the value of the stimulus intensity and the participant's response represent a stimulus-response pair and constitute one trial. Stimulus-response pairs were recorded, and 90 trials were conducted for each stimulus range tested.

Tests were run on a total of 15 participants, 8 females and 7 males, whose ages ranged from 20 to 35 years with a mean age of 23.5 years. Each participant was tested for a number of sound intensity ranges, R, at each frequency. For example, at 1000 Hz all participants were tested for the following five ranges: 1 to 9, 1 to 36, 1 to 54, 1 to 72, and 1 to 90 dB HL, while at 125 Hz the same participants were tested at ranges of 1 to 9, 1 to 27, 1 to 45, 1 to 63, and 1 to 81 dB HL. Our selection of a set of ranges at a given frequency was constrained to the stimulation limits of our audiometer, set within safety standards. The six stimulus frequencies tested in this study were 125, 250, 1000, 1500, 4000, and 8000 Hz.

All experiments were conducted in accordance with a protocol authorized by the University of Toronto Research Ethics Board. All participants gave informed consent before their inclusion in the study.

Analytical Methods

In a given trial, a participant will occasionally identify the intensity of the stimulus tone correctly. More often than not, participants will produce errors in identification. In keeping with Thurstone's conjecture (1927) and the findings of others (Durlach and Braida, 1969; Luce et al., 1976), Norwich et al. (1998) observed that a participant's response tends to be normally distributed about the stimulus.

Using the collection of stimulus-response pairs obtained in an experimental session, one can construct the normal distribution that underlies the error in a participant's responses for a specified range. On a given trial, let the stimulus intensity be x dB and the participant's response be y dB. By subtracting the stimulus from the response, one obtains $w = y - x$. For example, $w = 0$ implies that the participant provided a correct response, and $w = -3$ indicates that the participant provided an estimate that was 3 dB less than the actual stimulus. Hence, each trial results in a single value for w, and the collection of 90 such values can be compiled into a histogram as illustrated in Fig. 1, taken from participant F4 over the range 1 to 9 dB tested at 250 Hz. As one can see, the distribution fits well to a normal distribution centered at $w = 0$ and has standard deviation $s = 1.857$ dB.

Note that at a fixed frequency standard deviation, s, remains fixed for a given stimulus range R. However, s tends to increase as R is increased. This observation was reported by us and other investigators (Durlach and Braida, 1969; Luce et al, 1976; Norwich et al. 1998), and confirmed in this study. The power function exponent, n, can be estimated in the following manner.

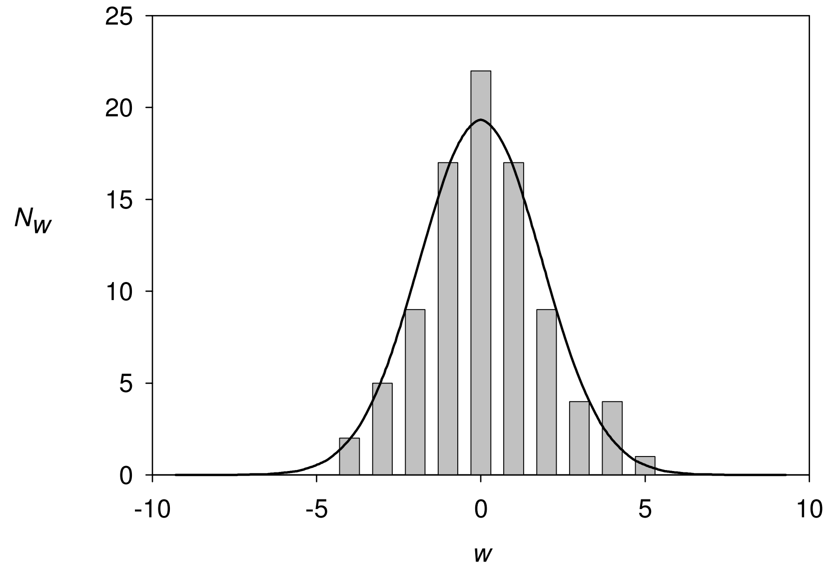


Figure 1. Resultant histogram obtained from Participant F4 over the range 1 to 9 dB HL at 250 Hz. The smooth curve is a normal distribution with a mean of zero and the same variance as the histogram. It represents approximately the distribution that underlies the subject's responses.

For a given auditory frequency, we plot s, as calculated from a given set of normally distributed data, against R. For example, at 1000 Hz, for participant F2 over the stimulus ranges 1 to 9, 1 to 36, 1 to 54, 1 to 72, and 1 to 90 dB HL, the plot of s against R is shown in Fig. 2. A least squares regression line is fitted to the data using a weighting factor equal to the inverse square of s (cf. Snedecor and Cochrane 1967). The regression line can be expressed in the form

$$s = aR + b \,. \tag{2}$$

Equation (2) is similar in form to the equation proposed by Durlach and Braida (1969).

We now propose a method for obtaining the loudness exponent, n, from the increase of s with R, as defined by Equation (2). The first step is to relate the stimulus range, R, to a mean stimulus value I. Because stimulus intensities are selected from the range using a uniform distribution, the mean intensity value presented to the participant for a range R is equal to half the stimulus range. That is,

$$I = R/2 \,. \tag{3}$$

Substituting Equation (3) into Equation (2), we find that

$$s = 2aI + b \,. \tag{4}$$

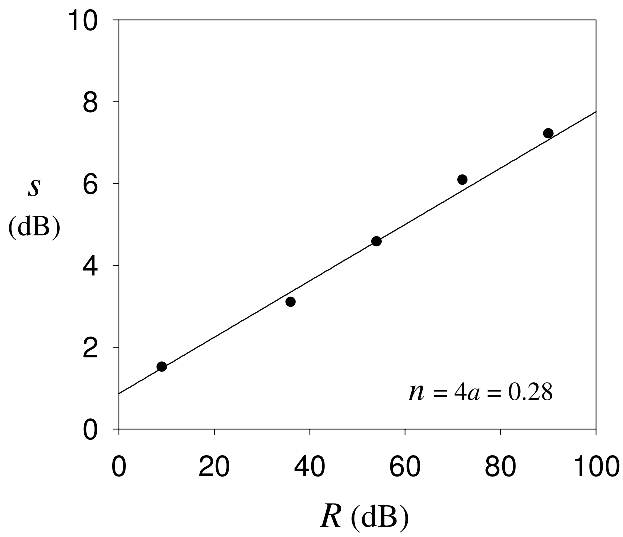


Figure 2. Graph of standard deviation, s, versus range, R, for Participant F2 at 1000 Hz. The regression line is weighted by $1/s^2$.

Recall that s and I are logarithmic quantities expressed in decibel measure. If we let σ be the linear measure corresponding to s, and φ be the linear measure corresponding to I, then it can be shown that Equation (4) yields $\sigma \propto \varphi^{2a}$. By squaring both sides, we find that

$$\sigma^2 \propto \varphi^{4a}. \tag{5}$$

Stevens' power law can be written as

$$\psi \propto \varphi^n. \tag{6}$$

Since a and n are constants, we may define the constant $m = 4a/n$.

Then

$$\psi^m \propto \varphi^{n\,m} = \varphi^{4a} \tag{7}$$

From (5) and (7), σ^2 / ψ^m = constant. In the ensuing statistical analysis, we let $m = 1$, and hence $n = 4a$.

Statistical Analysis and Results

We first conducted an analysis of variance (ANOVA) of exponents between frequencies within each gender. The null hypothesis, H_0, is that the mean n-values for each frequency (in Hz: 125, 250, 1000, 1500, 4000 and 8000) are equal. For both males and females, we were unable to reject this null hypothesis, $p > 0.05$. Since intragender exponents did not differ significantly between frequencies, we were justified in pooling exponents across all frequencies within each gender.

Table 1

Average exponents pooled across all frequencies (125, 250, 1000, 1500, 4000, 8000 Hz) for females and males

	avg n	sd	# of n's	t-test
females	0.3151	0.0745	34	reject H_0
males	0.2148	0.0518	22	$p < 0.0001$

We then utilized the pooled variance t-test (Smith 1998), which makes use of a pooled or averaged sample variance for the two groups together. In order to employ this method it is first necessary to ascertain that population variances for each group, males and females, are equal. Therefore, preliminary to the t-test, we conducted an F-test on the ratio of the sample variances of the two groups. Here, H_0 is that the two sample variances are equal. We failed to reject H_0, $p > 0.05$. Therefore, we were permitted to calculate a pooled variance for the two groups, and to proceed with the pooled variance t-test for differences between mean exponents of males and females.

We next compared exponents combined across frequency for each group. That is, we compared 34 exponents obtained from 8 females and 22 exponents obtained from 7 males. Results from this pooled variance t-test are shown in Table 1 and indicate that female exponents are significantly higher than male exponents, $p < 0.0001$.

We can summarize the results as follows. *Within a gender*, no significant difference between values of the exponent, n, could be measured across stimulus frequency. However, *between genders*, there is a significant difference in the values of n. In grouping all frequencies together, the difference between the genders was significant statistically: values of n for female participants exceeded those for male participants, $p < 0.0001$.

Discussion

When participants were required to identify the intensity of tones to the nearest decibel, it was found that errors were distributed in accordance with the normal distribution. Variance was found to increase with range; a plot of standard deviation versus range yields a straight line with slope, a. Four times a can be identified with the loudness exponent. Statistically significant differences are seen between measures of loudness exponents (obtained as $4a$) for males and females. These findings accord with the inferences of McGuinness (1974).

One possible implication of the gender difference in loudness exponents is that females experience a broader range of loudness than their male counterparts with similar thresholds. Alternatively phrased, for a fixed range of intensities, a steeper loudness function implies that females are more sensitive to a given physical range of tones than males. An interesting finding that emerged from our analysis is that errors in tone identification can be a measure of loudness. That is, ψ^m, which is a *measure* of loudness, is proportional to φ^{4a} (Equation (7)) which by (5) is a measure of error in tone identification. More studies are needed and are underway to further explore the results found here.

Acknowledgements

This research was funded through a Discovery Grant to K. H. Norwich from the Natural Sciences and Engineering Research Council of Canada.

References

Don, M., Ponton, C. W., Eggermont, J. J. & Masuda, A. (1993). Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response. *Journal of the Acoustical Society of America, 94*, 2135-2148.

Durlach, N. I. & Braida, L. D. (1969). Intensity Perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46*, 372-383.

Luce, R. D., Green, D. M. & Weber, D. L. (1976). Attention bands in absolute identification. *Perception & Psychophysics, 20*, 49-54.

McFadden, D. (1998). Sex differences in the auditory system. *Developmental Neuropsychology, 14*, 261-298.

McGuinness, D. (1974). Equating individual differences for auditory input. *Psychophysiology, 11*, 113-120.

Mori, S. & Ward, L. (1995). Pure feedback effects in absolute identification. *Perception & Psychophysics, 7*, 1065-1079.

Norwich, K. H., Wong, W. & Sagi, E. (1998). Range as a factor determining the information of loudness judgments: Overcoming small sample bias. *Canadian Journal of Experimental Psychology, 52*, 63-70.

Smith, P. J. (1998). *Into Statistics: A guide to understanding statistical concepts in engineering and the sciences*. Springer-Verlag, Singapore.

Snedecor, G. W. & Cochrane, W. G. (1967). *Statistical methods, 6th Ed'n*. The Iowa State University Press, Ames, IA.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review, 34*, 273-286.

EFFECTS OF DOPAMINERGIC THERAPY ON TIME PERCEPTION IN PARKINSON'S DISEASE: AN EVENT-RELATED FMRI STUDY

Petr Dušek[1], Robert Jech[1], Ji í Wackermann[2] and Josef Vymazal[3]

[1]Department of Neurology, 1st Medical Faculty of Charles University, Kate inská 30, 120 00 Prague, Czech Republic; [2]Institute for Frontier Areas of Psychology, Wilhelmstr. 3a, 79098 Freiburg i. Br., Germany; [3]Na Homolce Hospital, Roentgenova 37/2, 150 00 Prague, Czech Republic
e-mail: dusek@nanoware.cz

Abstract

Parkinson's disease patients (PD) are deficient in time perception. We investigated the pathophysiology of this deficit using fMRI.

A Time reproduction task (TRT) was employed - subjects had to retain the duration of visually presented stimuli and reproduce it by pressing a button. Patients were examined twice: on dopaminergic medication (ON state), and after withdrawal from medication (OFF state). Brain activity related to the button pressing was visualized for both states.

In the OFF state, the parietal, dorsolateral prefrontal (DLPFC), opercular and cingulate cortices were activated bilaterally. The supplementary motor area (SMA) and putamen were activated in the right hemisphere.

In the ON state, activation was broadly reduced. Activation of the frontostriatal network and opercular cortex disappeared in the left hemisphere. Activations of the right hemisphere's SMA and DLPFC were reduced.

Our results suggest aberrant activation or compensatory engagement of homologous brain areas in PD patients OFF medication.

It has been reported, that Parkinson's disease (PD) patients are deficient in the estimation of time in the range of milliseconds (Artieda et al., 1992) as well as seconds (Pastor et al., 1992). This deficit was claimed to be improved after L-DOPA replacement, deep brain stimulation (DBS) (Koch et al., 2004a) and even repetitive transcranial magnetic stimulation (rTMS) over the right dorsolateral prefrontal cortex (DLPFC) (Koch et al., 2004b).

Functional imaging research of time perception in healthy volunteers has shown that brain areas of the brain activated during processing of subsecond and suprasecond intervals are not identical. Whereas subsecond intervals seems to be processed by sensorimotor cortex and supplementary motor area (SMA), suprasecond intervals more likely activate the DLPFC and associative parietal areas (Lewis et al., 2003).

In PD patients, dysfunction in all the above mentioned areas has been previously documented (Grafton,2004; Haslinger et al., 2001). However, little is known about whether and how these dysfunctions contribute to the deficit in time perception. Functional magnetic resonance imaging (fMRI) study in PD patients with a finger tapping task at a frequency of 600ms showed decreased activation of sensorimotor cortex and medial premotor system including the SMA, as would be expected when subsecond intervals were employed (Elsinger et al., 2003).

Up until now, no study has investigated brain areas activated during time perception of suprasecond intervals in PD patients. In our study, we used fMRI to investigate brain activity during time reproduction task (TRT) of various intervals in the suprasecond range in PD patients in on medication (ON) and after medication withdrawal (OFF) states. We expected aberrant activation in DLPFC and parietal cortex in the OFF state, which would be normalized after replacement of dopaminergic therapy.

Table 1: clinical characteristics of PD patients

patient	major hemibody involvement	age (years)	disease duration (years)	L-DOPA equivalent (mg)	UPDRS III score	
					OFF	ON
1	right	55	10	700	27.5	6
2	left	45	8	725	27.5	19,5
3	right	63	6	550	31.5	28
4	right	65	2	475	30	25,5
5	left	48	8	550	32.5	16
6	right	70	9	1080	27.5	13
7	right	64	4	600	17	12,5
8	left	62	1	0	16.5	11

Method

Eight mild to moderate PD patients (all males, strongly right-handed, average age 59 years, average disease duration 6 years) participated in our study. All of our subjects were asymmetrically affected with parkinsonian symptoms, five with major right hemibody and three with left hemibody involvement. PD symptoms were primarily akinetic-rigid. Except one, all subjects were on stable dopaminergic medication, which consisted of L-DOPA/carbidopa and dopamine agonist.

The TRT consisted of two phases: encoding and reproduction. During the encoding phase, a visual stimulus (a gray square with a red cross in the middle) of eight various predefined duration was presented to the subjects. The durations were: 5.00, 5.95, 7.07, 8.41, 10.00, 11.89, 14.14, and 16.82 seconds (i.e., a geometric sequence $x_{i+1} = x_i \,.\, 2^{1/4}$). These durations were shown pseudo randomly, each of them twice during one session. After the encoding phase, subjects had to retain the presented duration and reproduce it by pressing a joystick pushbutton. The reproduction phase started after an interstimulus interval of 10 seconds with the appearance of a grey square with a green cross in the middle. This visual stimulus remained on for 150 % of the duration of the encoding interval, regardless as to whether the subject pressed the pushbutton or not. The next interval reproduction followed after 10 seconds. The duration of the whole task was 12.5 minutes. Subjects were instructed to avoid the strategy of mental counting.

The task was performed twice by each PD patient in the ON and OFF states, with the interval between sessions ranging from 1 to 5 weeks. In the OFF state, the dopamine agonist was discontinued for a minimum of 3 days and L-DOPA for minimum of 12 hours prior to scanning. One subject, who did not take any regular medication, was given an extraordinary dose of L-DOPA (375mg?) prior to scanning in the ON state. In the ON state, subjects were asked to take regular medication. The order of the ON and OFF sessions was counterbalanced. Each MRI session was preceded by TRT practice whilst off the MR scanner and by Unified Parkinson's Disease Rating Scale (UPDRS) III examination. Prior the fMRI session, each patient was tested using a behavioral version of the TRT task.. The task consisted only three durations (5, 10 and 16.82 s,) „each being used seven times during the task. Two parameters were evaluated, mean reproduction time as a measure for time estimation accuracy and standard deviation as a measure for response variability. Multivariate analysis of variance with repetition was used for data analysis.

The fMRI was performed on a Siemens Symphony 1.5 T scanner using the GE-EP sequence (TR=2.9 s, TE=56 ms, FA=90°). 260 dynamic volumes consisting of 24 axial sections, each 4 mm thick, were recorded during the session.

The event-related fMRI data analysis was performed to search for brain areas related to the joystick-button press event. Data was preprocessed and analyzed using the SPM2 software (Wellcome Dept. of Cognitive Neurology, London, UK). Motion artifacts were corrected using the least squares method. Differences in slice time acquisition were corrected; all scans were spatially normalized into

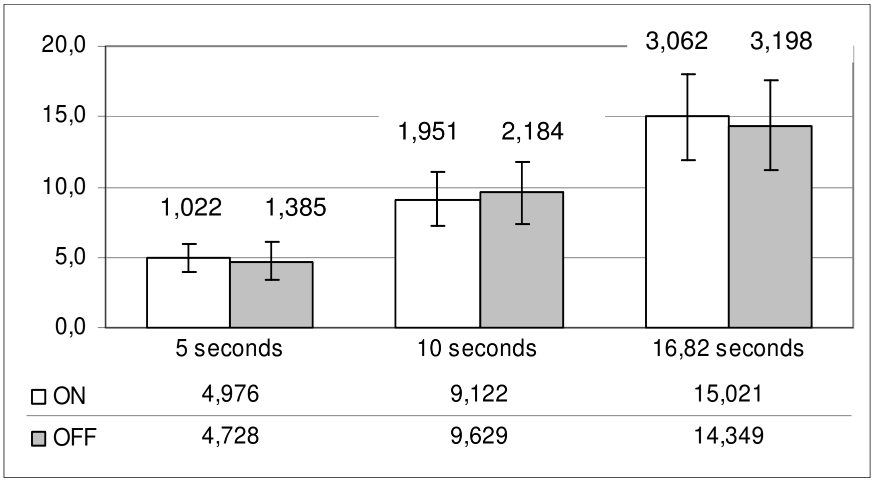


Figure 1: mean time reproduction responses for 5, 10 and 16.82 second intervals for ON and OFF states. Standard deviations are depicted as the error bars with values above them.

standardized stereotactic space (Montreal Neurological Institute coordinates as provided by SPM2) and smoothed with a 10-mm Gaussian filter to suppress residual inter-individual differences. Each event was convolved with the expected hemodynamic response. Results of the 2nd level (random effects) analysis were obtained by separate analysis of the ON and OFF states using a one sample t-test. A paired t-test was applied for comparison between both medication states. Results were thresholded at the $P<0.001$ level uncorrected for multiple comparisons.

Results

In all of our subjects motor symptoms were improved after dopaminergic therapy. The mean UPDRS III score dropped from 26.3 (± 6.16) in the OFF state to 16.4 (± 7.46) in the ON state. The behavioral data from the practice task showed that PD patients were equally accurate in time reproductions in both states, but slightly greater variance of responses in the OFF state was observed. However, it did not reach the level of statistical significance (Figure 1).

In the OFF state, the event-related fMRI analysis related to the pushbutton press event showed large activation throughout the whole brain. As shown in the table 2, the activation was seen in the parietal cortex, DLPFC and cingulate cortex bilaterally. The supplementary motor area (SMA), putamen, superior temporal and inferior lateral prefrontal cortices were activated in the right hemisphere only. The insular cortex was activated in the left hemisphere only.

In the ON state, the activation pattern was broadly reduced. In the ON state, only 600 voxels survived thresholding compared to 2140 voxels in the OFF state. In the left hemisphere, activation of parietal cortex, DLPFC and insular cortex disappeared. Compared to the OFF state, activation in the left primary motor cortex (M1) appeared in the left hemisphere. In the right hemisphere, activation of SMA was less apparent and activation of putamen disappeared. Moreover, the cluster of activation in the right DLPFC moved anteriorly. Large bilateral activation in anterior cingulate seen in the OFF state disappeared in the ON state as well.

When paired t-test was applied to compare both states, blob in the left DLPFC was active in the OFF compared to the ON state. Other areas previously mentioned were not observed using this analysis.

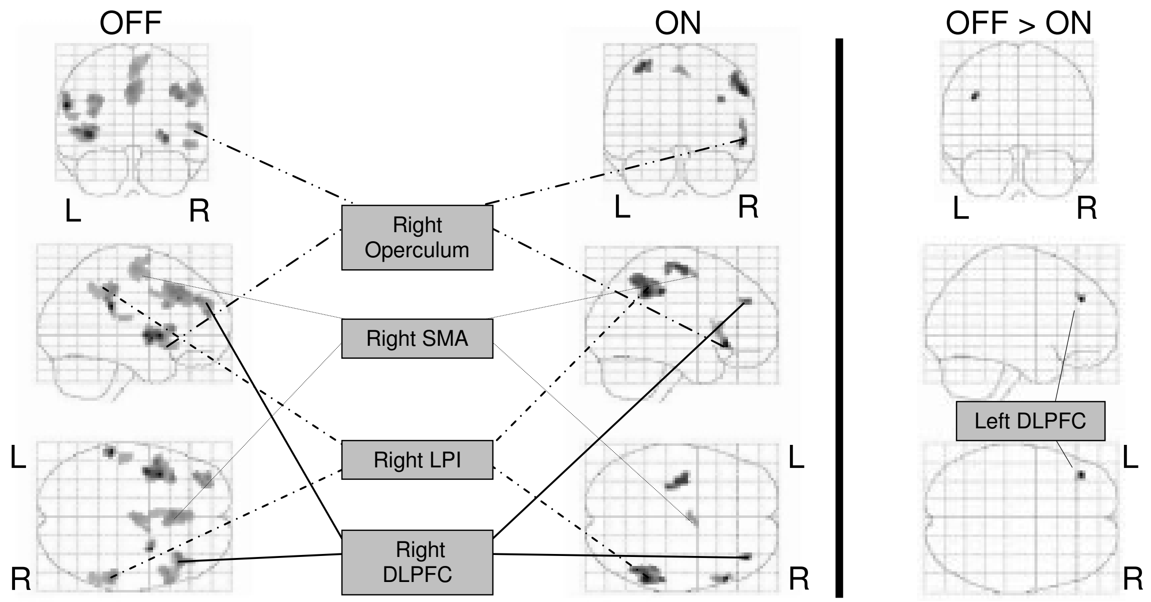


Figure 2: Group analysis in the OFF and ON states are shown on the left. Brain areas activated in both states are depicted. Paired analysis with area activated in the OFF state compared to the ON state is shown on the right (group analysis, P<0.001 uncorrected). SMA – Supplementary motor area, LPI – lobulus parietalis inferior, DLPFC – dorsolateral prefrontal cortex, M1 – primary motor cortex

Discussion

In the behavioral part of our study, there was not significant a difference between the ON and OFF states, either in the accuracy or in the variability of response. However, response variability seemed to be slightly increased in the OFF state.

However, time reproduction task with empty time intervals in PD patients did not gain significant results in previous study (Pastor et al., 1992).

Many of the brain areas activated in our study – the DLPFC, SMA, parietal cortex and putamen have been already suggested to be essential for time perception in healthy subjects (Rubia et al., 2004). Previous studies mainly stressed the importance of the right fronto-parietal network in time perception of suprasecond intervals (Harrington et al., 1998). However, activation of the left hemisphere has also been described by some authors (Rao et al., 2001; Rubia et al., 1998).

Compared to healthy subjects performing the same task (Jech et al., 2004), PD patients in the OFF state demonstrated additional activation in the left DLPFC, left parietal cortex, left insular cortex, bilateral anterior cingulate cortex and right putamen and no activation of the left M1. Moreover, the extent of activation in the right DLPFC was larger than seen in healthy subjects.

Abnormally elevated activation of DLPFC and parietal cortex, which decreased after L-DOPA, was previously described in PD patients during performing of Tower of London task (Cools et al., 2002) and N-Back task (Mattay et al., 2002). During motor sequence learning (Nakamura et al., 2001) and sentence processing (Grossman et al., 2003), bilateral DLPFC activation was even described, similarly to our study. During TRT of suprasecond intervals, encoded intervals must be retained in working memory and decision must be taken before pressing the pushbutton. It is certainly a complex cognitive task. Enhanced prefrontal activation seems to be a typical finding in executive tasks in PD patients OFF medication. Two possible explanations exist for this phenomenon: it may reflect inefficiently broadened activation due to a loss of dopaminergic mesocortical focusing function in the DLPFC, or it may reflect adaptive compensatory activation of homologous brain areas. Authors of a study, which showed that PD patients performing sequence learning task at the same level as healthy subjects had enhanced brain activation, argue that this was due to activation of compensatory brain resources (Carbon et al., 2003).

Table 2: regions activated at the event of the button press
Results for the ON state, OFF state and comparison of the OFF relative to the ON state are shown.
(BA – Brodmann area; x, y, z – coordinates in standardized stereotactic space, T value)
Areas exclusively activated in the OFF or ON states are **BOLD**

	region	BA	voxels	x	y	z	T
• *OFF state:*							
Left inferior parietal lobule		40	104	-58	-34	26	15.24
Left insular cortex			477	-38	8	0	14.27
Left superior frontal gyrus	DLPFC	9/10	189	-36	42	32	8.42
Bilat. anterior cingulate		32	448	0	22	40	8.61
Right superior frontal gyrus	DLPFC	9	227	42	26	34	10.02
Right putamen			57	30	2	-2	12.46
Right inferior parietal lobule		40	265	56	-34	40	8.20
Right inferior frontal gyrus		45	84	58	18	6	7.38
Right medial frontal gyrus	SMA	6	199	2	-12	60	7.01
Right superior temporal gyrus	PTC	38	91	50	14	-10	6.87
• *ON state:*							
Left precentral gyrus	M1	4	125	-32	-18	64	6.45
Right inferior frontal gyrus		47	73	56	26	-8	6.92
Right inferior parietal lobule		40	338	56	-46	40	7.02
Right superior frontal gyrus	DLPFC	9/46	37	36	44	32	6.19
Right medial frontal gyrus	SMA	32	27	-2	-8	60	5.34
• *OFF > ON state*							
Left superior frontal gyrus	DLPFC	9	15	-38	40	34	3.64

After dopaminergic replacement, the activation of the left fronto-parietal network disappeared. Similar findings were reported in all above-mentioned studies. Again, the effect of dopamine might have caused increase of efficiency of information processing in cerebral cortex causing the compensatory activation no longer to be necessary. It might also have ameliorated the deficit of the mesocortical dopaminergic system, which potentially caused aberrant activation.

The only area which survived statistical comparison between the OFF and ON states, was the left DLPFC, suggesting it is the most robust finding of our study. Regarding this is only a pilot study, more areas activated may get statistically significant in statistical comparison of both states with an increasing number of subjects.

The primary motor cortex was activated only after replacement of dopaminergic therapy. This finding is in agreement with results of study using finger-tapping task in PD patients (Elsinger et al., 2003). We assume this activation to be related to motor aspects of the task. In our previous study with healthy subjects, most of left M1 activation disappeared after motion subtraction (Jech et al., 2004). The activation of M1 might thus be a physiological correlate of improved motor symptoms of PD patients.

The importance of the basal ganglia in time perception was stressed in many studies (Rubia et al., 2004). However, why the right putamen was activated in the OFF state is unclear. The above-mentioned study (Elsinger et al., 2003) showed opposite pattern of putamen activation. Contrary, microrecording from basal ganglia generally showed increased neuronal firing rate during the OFF state (Dostrovsky et al., 2000), which might be the explanation for hyperactivation seen in our study.

Our study is the first to show physiological underpinnings of PD patients' deficit in time perception. The results show similar pattern of activation as other studies investigating executive dysfunction in PD patients. This dysfunction is not specific for time perception and reflects more general impairment, which can be called "frontal like syndrome".

References

Artieda J, Pastor MA, Lacruz F, Obeso JA (1992). Temporal discrimination is abnormal in Parkinson's disease. Brain 115 Pt 1: 199-210

Carbon M, Marie RM (2003). Functional imaging of cognition in Parkinson's disease. Curr Opin Neurol 16: 475-480

Cools R, Stefanova E, Barker RA, Robbins TW, Owen AM (2002). Dopaminergic modulation of high-level cognition in Parkinson's disease: the role of the prefrontal cortex revealed by PET. Brain 125: 584-594

Dostrovsky JO, Levy R, Wu JP, Hutchison WD, Tasker RR, Lozano AM (2000). Microstimulation-induced inhibition of neuronal firing in human globus pallidus. J Neurophysiol 84: 570-574

Elsinger CL, Rao SM, Zimbelman JL, Reynolds NC, Blindauer KA, Hoffmann RG (2003). Neural basis for impaired time reproduction in Parkinson's disease: an fMRI study. J Int Neuropsychol Soc 9: 1088-1098

Grafton ST (2004). Contributions of functional imaging to understanding parkinsonian symptoms. Curr Opin Neurobiol 14: 715-719

Grossman M, Cooke A, DeVita C, Lee C, Alsop D, Detre J, Gee J, Chen W, Stern MB, Hurtig HI (2003). Grammatical and resource components of sentence processing in Parkinson's disease: an fMRI study. Neurology 60: 775-781

Harrington DL, Haaland KY, Knight RT (1998). Cortical networks underlying mechanisms of time perception. J Neurosci 18: 1085-1095

Haslinger B, Erhard P, Kampfe N, Boecker H, Rummeny E, Schwaiger M, Conrad B, Ceballos-Baumann AO (2001). Event-related functional magnetic resonance imaging in Parkinson's disease before and after levodopa. Brain 124: 558-570

Jech R, Dušek P, Wackermann J (2004). FMRI correlates of time perception. Fechner Day Proceedings: 158-163

Koch G, Brusa L, Caltagirone C, Oliveri M, Peppe A, Tiraboschi P, Stanzione P (2004a). Subthalamic deep brain stimulation improves time perception in Parkinson's disease. Neuroreport 15: 1071-1073

Koch G, Oliveri M, Brusa L, Stanzione P, Torriero S, Caltagirone C (2004b). High-frequency rTMS improves time perception in Parkinson disease. Neurology 63: 2405-2406

Lewis PA, Miall RC (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. Curr Opin Neurobiol 13: 250-255

Mattay VS, Tessitore A, Callicott JH, Bertolino A, Goldberg TE, Chase TN, Hyde TM, Weinberger DR (2002). Dopaminergic modulation of cortical function in patients with Parkinson's disease. Ann Neurol 51: 156-164

Nakamura T, Ghilardi MF, Mentis M, Dhawan V, Fukuda M, Hacking A, Moeller JR, Ghez C, Eidelberg D (2001). Functional networks in motor sequence learning: abnormal topographies in Parkinson's disease. Hum Brain Mapp 12: 42-60

Pastor MA, Artieda J, Jahanshahi M, Obeso JA (1992). Time estimation and reproduction is abnormal in Parkinson's disease. Brain 115 Pt 1: 211-225

Rao SM, Mayer AR, Harrington DL (2001). The evolution of brain activation during temporal processing. Nat Neurosci 4: 317-323

Rubia K, Overmeyer S, Taylor E, Brammer M, Williams S, Simmons A, Andrew C, Bullmore E (1998). Prefrontal involvement in "temporal bridging" and timing movement. Neuropsychologia 36: 1283-1293

Rubia K, Smith A (2004). The neural correlates of cognitive time management: a review. Acta Neurobiol Exp (Wars) 64: 329-340

FROM FECHNER TO FECHNER, AND BEYOND

Ehtibar N. Dzhafarov
Purdue University (ehtibar@purdue.edu)

Hans Colonius
Oldenburg University (hans.colonius@uni-oldenburg.de)

Abstract

Dzhafarov and Colonius's theory of Fechnerian Scaling deals with arbitrary spaces of stimuli endowed with probability-of-different functions $\psi(\mathbf{x},\mathbf{y})$. The cornerstone of this theory is the Law of Regular Minimality, according to which $h(\mathbf{x}) = \arg\min_{\mathbf{y}} \psi(\mathbf{x},\mathbf{y})$ and $g(\mathbf{y}) = \arg\min_{\mathbf{x}} \psi(\mathbf{x},\mathbf{y})$ are well-defined functions, and $g \equiv h^{-1}$. Due to this law the psychometric increments $\Psi^{(1)}(\mathbf{x},\mathbf{y}) = \psi(\mathbf{x},\mathbf{y}) - \psi(\mathbf{x}, h(\mathbf{x}))$ and $\Psi^{(2)}(\mathbf{y},\mathbf{x}) = \psi(\mathbf{x},\mathbf{y}) - \psi(g(\mathbf{y}),\mathbf{y})$ can be cumulated along "allowable paths" within stimulus spaces yielding the paths' lengths. The Fechnerian distance $G(\mathbf{a},\mathbf{b})$ between $\mathbf{a}$ and $\mathbf{b}$ is defined as the infimum of the lengths of all allowable closed loops from $\mathbf{a}$ to $\mathbf{b}$ and back. The notion of an allowable loop, however, has been defined differently for continuous spaces (piecewise smooth arcs) and discrete spaces (finite chains). In this paper we introduce a new mathematical notion, Dissimilarity Function, and assume that $\Psi^{(1)}$ and $\Psi^{(2)}$ are dissimilarity functions. Allowable loops in all stimulus spaces in this new version of the theory can be defined as finite cycles of stimuli, $G(\mathbf{a},\mathbf{b})$ being the infimum of the sums of the chained psychometric increments along all such cycles containing $\mathbf{a}$ and $\mathbf{b}$. This construction includes previous developments of Dzhafarov-Colonius's theory as special cases, but it rather radically departs from the spirit of Fechner's original theory.

In Fechner's original theory every point on a unidimensional stimulus continuum is assigned a measure of its dissimilarity from its "immediate neighbors," and this dissimilarity measure is integrated over the interval between two stimuli to yield the subjective distance between them. Since Fechner assumed that the unidimensional continuum of stimuli monotonically corresponds to a unidimensional continuum of "sensations," the local dissimilarity measure in this theory can be derived from greater-less judgments, e.g., by computing the slopes of the probability-of-greater functions at their medians or, as a crude approximation, the reciprocals of JNDs. In Dzhafarov (2001) and Dzhafarov & Colonius (1999a) we argue that the integration of a local dissimilarity measure constitutes the essence of Fechner's theory. Departing from this idea (the "*from Fechner*" of the title) we developed a theory we called Multidimensional Fechnerian Scaling, MDFS (Dzhafarov, 2002a-d, 2004; Dzhafarov & Colonius, 1999a, b, 2001): unidimensional stimulus continua are replaced in MDFS with arbitrary open regions of n-dimensional real-valued vectors, the greater-less judgments are replaced with same-different ones, and the dissimilarity measure (*psychometric increments*) is computed from the probability-of-different functions. When this dissimilarity measure is integrated along a piecewise smooth (i.e., continuously differentiable) arc connecting two stimuli, one gets this arc's length (we call it *psychometric length*); and the infimum of all such lengths across all piecewise smooth closed loops containing given two stimuli yields the subjective (Fechnerian) distance between these stimuli.

In Dzhafarov & Colonius (2005a) this theory was generalized to arbitrary spaces in which one can define piecewise smooth arcs. Unlike in MDFS, in this generalized theory all topological, geometrical, analytical, and metric properties of stimulus spaces are derived solely from

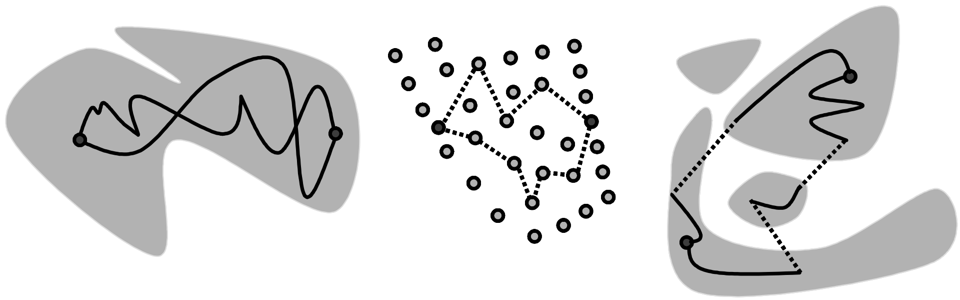


Figure 1: Allowable closed loops in (from left to right) continuous, discrete, and hybrid (discrete-continuous) stimulus spaces.

the properties of discrimination probabilities, with no reference to physical properties of stimuli. Therefore we call this theory *purely psychological* ("psychophysics without physics"). As a special case, one can return to Fechner's original theory ("*to Fechner*") and reformulate it with no reference to the representability of stimuli by real numbers, their Euclidean topology, order, additivity, or any other physical property explicitly or implicitly used, e.g., in the derivation of the celebrated logarithmic law. Physical measurements are only used as stimulus labels, each label exhaustively characterized by its discrimination probabilities from other labels (with regards to its *observation area*, as explained below); the Fechnerian distance between stimuli labeled $\mathbf{a}$ and $\mathbf{b}$ will remain precisely the same if the unidimensional continuum of stimuli is bijectively mapped, say, in a unit square.

[To prevent confusion: the "purely psychological" does not mean that one cannot or should not use physical measures, it only means that the Fechnerian distances among stimuli are logically and computationally independent of the choice of physical measures. Nothing prevents one, once Fechnerian distances have been computed, to ask whether the Fechnerian distance between stimulus and absolute threshold is, say, a power function of a specific physical measure of stimulus.]

In parallel to the Generalized Fechnerian Scaling for continuous spaces we developed another purely psychological theory, Fechnerian Scaling of Discrete Object Sets, FSDOS (Dzhafarov & Colonius, 2005b, in press a, c). Discrete object (or stimulus) sets are those in which the dissimilarity of any stimulus from all other stimuli cannot fall below some positive quantity. No two distinct stimuli in a discrete space can be connected by an arc, and the notion of a piecewise smooth closed loop here is replaced with that of a closed finite chain of stimuli (leading from $\mathbf{a}$ to $\mathbf{b}$ and back). Each link of such a chain is characterized by the dissimilarity of its endpoints, and the psychometric length is simply the sum of the dissimilarity values of its links. Although we do call these computations Fechnerian and have established useful relations between them and the continuous Fechnerian computations (Dzhafarov & Colonius, in press a, c), there is a conceptual gap between them, especially apparent when we consider hybrid spaces consisting of isolated continuous components (Dzhafarov & Colonius, 2005b): an allowable path here consists of piecewise smooth arcs within components and discrete jumps between them (Fig. 1). This situation seems undesirable as we do not have a clear operational boundary between continuous and discrete spaces. One might even argue that every discrete space (e.g., that of the written letters of an alphabet) is only an approximation to, or a categorization of, a continuous one (graphical modifications of written letters).

The revision we propose is seemingly simple: while the notion of an arc is not defined for discrete spaces, the notion of a finite chain is defined for all stimulus spaces without exception; the proposal therefore is to define the subjective distance between $\mathbf{a}$ and $\mathbf{b}$ as the

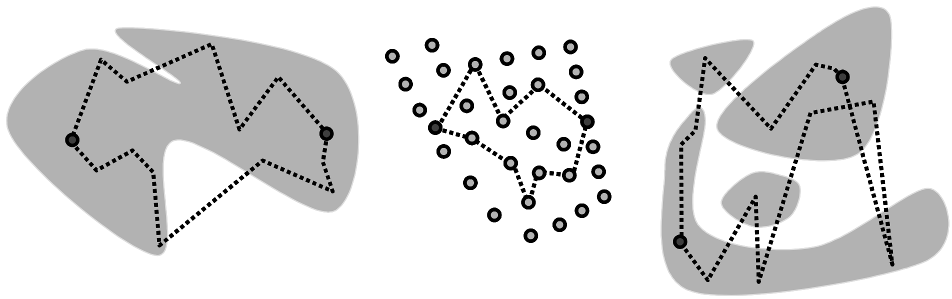


Figure 2: In Universal Fechnerian Scaling the allowable closed loops in all possible stimulus spaces are finite chains of stimuli.

infimum of psychometric length for all closed finite chains containing **a** and **b**. In other words, the computations in all possible stimulus spaces are the same as in FSDOS (see Fig. 2). The implementation of this simple idea, however, requires thorough rethinking of the basic concepts upon which we base our theory. In the following we present outlines of this new development (necessarily schematic due to the lack of space).

In essence, what we propose amounts to elevating FSDOS from somewhat of an afterthought it has been to becoming the very core of Fechnerian Scaling. One might challenge the continued use of the adjective "Fechnerian" for this theory: Fechner's idea of the integration of infinitesimally small dissimilarities is being replaced here, on the most basic level, by summation of finite dissimilarities along finite stimulus chains (although in continuous spaces, as explained below, this may lead to arcs and integration "in the limit," if the number of links in the chains whose lengths approach the infimum increases beyond bounds). In other words, with reference to the title, it may very well be the case that the present development leads us "*beyond Fechner*." At this stage, however, we prefer to retain the "Fechnerian" terminology for continuity with our previous work: the new version of the theory can be called *Universal Fechnerian Scaling*, to emphasize that one and the same construction is applied to all possible stimulus spaces.

Dissimilarity Function: A New Mathematical Notion

Let $\mathfrak{S}$ be a set with elements (stimuli) denoted by boldface lowercase symbols $(\mathbf{a}, \mathbf{b}, \mathbf{x}, \mathbf{y}, ...)$. A string of k such symbols, $\mathbf{x}_1...\mathbf{x}_k$, represents a stimulus chain of cardinality k. We use the notation of the form $[prefix]\mathbf{x}_1...\mathbf{x}_k$ to denote various numerical characteristics of chain $\mathbf{x}_1...\mathbf{x}_k$ (e.g., $D\mathbf{ab}$, $D\mathbf{x}_1...\mathbf{x}_k$, $\psi\mathbf{xy}$, ...).

Consider a function $D : \mathfrak{S} \times \mathfrak{S} \to [0, 1]$ (in the last section we will show how to generalize this to mapping into $\mathbb{R}^+$). We call D a *deviation function* if

$$D\mathbf{ab} = 0 \Longleftrightarrow \mathbf{a} = \mathbf{b}. \tag{1}$$

A deviation function D is called a *dissimilarity function* if it has the following two properties:

$$\left(D\mathbf{a}'_n\mathbf{a}_n \to 0\right) \wedge \left(D\mathbf{b}'_n\mathbf{b}_n \to 0\right) \Longrightarrow D\mathbf{a}'_n\mathbf{b}'_n - D\mathbf{a}_n\mathbf{b}_n \to 0, \tag{2}$$

and, for any sequence of chains $\left\{\mathbf{a}\mathbf{x}_1^n...\mathbf{x}_{k_n}^n\right\}_{n\in\mathbb{N}}$,

$$D\mathbf{a}\mathbf{x}_1^n...\mathbf{x}_{k_n}^n \to 0 \Longrightarrow D\mathbf{a}\mathbf{x}_{k_n}^n \to 0, \tag{3}$$

where $D\mathbf{x}_1...\mathbf{x}_k$ is defined as $\sum_{i=1}^{k-1} D\mathbf{x}_i\mathbf{x}_{i+1}$. Note that if D is a metric, then (1), (2), and (3) are satisfied trivially.

A dissimilarity function D induces a *metric* $G : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}^+$ by means of the following construction:

$$G\mathbf{ab} = \inf_{\{\mathbf{a},\mathbf{b}\}\subseteq\{\mathbf{x}_1,\ldots,\mathbf{x}_k\},k\in\mathbb{N}} D\mathbf{x}_1...\mathbf{x}_k\mathbf{x}_1 = \inf_{\mathbf{y}_1...\mathbf{y}_{k'},k'\in\mathbb{N}} D\mathbf{ay}_1...\mathbf{y}_{k'}\mathbf{b} + \inf_{\mathbf{z}_1...\mathbf{z}_{k''},k''\in\mathbb{N}} D\mathbf{bz}_1...\mathbf{z}_{k''}\mathbf{a}, \tag{4}$$

where the infimum on the left is taken over all closed loops $\mathbf{x}_1...\mathbf{x}_k\mathbf{x}_1$ in $\mathfrak{S}$ containing $\mathbf{a}$ and $\mathbf{b}$. The topologies based on open D-balls and open G-balls coincide. Generally, $G\mathbf{ab} \leq D\mathbf{ab} + D\mathbf{ab}$, but if D is an oriented metric (a metric which is not necessarily symmetrical), $G\mathbf{ab} = D\mathbf{ab} + D\mathbf{ab}$. Denoting by $\min k_{\text{inf}}(\mathbf{a}, \mathbf{b})$ the smallest cardinality k for which

$$\inf_{\mathbf{y}_1...\mathbf{y}_i,i\leq k} D\mathbf{ay}_1...\mathbf{y}_i\mathbf{b} = \inf_{\mathbf{y}_1...\mathbf{y}_i,i\in\mathbb{N}} D\mathbf{ay}_1...\mathbf{y}_i\mathbf{b},$$

we can have three cases: $\min k_{\text{inf}}(\mathbf{a}, \mathbf{b}) = 0$ (this will hold for all $\mathbf{a}, \mathbf{b}$ if D is an oriented metric), $0 < \min k_{\text{inf}}(\mathbf{a}, \mathbf{b}) < \infty$, and $\min k_{\text{inf}}(\mathbf{a}, \mathbf{b}) = \infty$. In the latter case, with some additional assumptions about $\mathfrak{S}$, $G\mathbf{ab}$ may coincide with the infimum of lengths of all closed arcs containing $\mathbf{a}, \mathbf{b}$.

Psychometric Increments as Dissimilarity Functions

We begin by briefly recapitulating the basic conceptual setup of our theory (Dzhafarov & Colonius, 2005a). A *discrimination system* is $(\mathfrak{S}_1^*, \mathfrak{S}_2^*, \psi^*)$ where $\mathfrak{S}_1^*, \mathfrak{S}_2^*$ are stimuli in, respectively, first and second *observation areas* (e.g., stimuli presented first and stimuli presented second); $\psi^* : \mathfrak{S}_1^* \times \mathfrak{S}_2^* \to [0, 1]$ is a probability-of-different function (see Dzhafarov & Colonius, 2005a, b and in press a,b, for a variety of meanings of the notions "observation area" and "same-different"). A *reduced discrimination system* is $\left(\mathfrak{S}_1, \mathfrak{S}_2, \tilde{\psi}\right)$, obtained from $(\mathfrak{S}_1^*, \mathfrak{S}_2^*, \psi^*)$ by lumping together *psychologically equal* stimuli in each of the observation areas: $\mathbf{a}$ and $\mathbf{b}$ in $\mathfrak{S}_1^*$ (or in $\mathfrak{S}_2^*$) are equal if $\psi^*\mathbf{ay} = \psi^*\mathbf{by}$ (resp., $\psi^*\mathbf{xa} = \psi^*\mathbf{xb}$) for all $\mathbf{y} \in \mathfrak{S}_2^*$ (resp., $\mathbf{x} \in \mathfrak{S}_1^*$).

The cornerstone of our theory is the *Law of Regular Minimality*: for any (reduced) discrimination system $\left(\mathfrak{S}_1, \mathfrak{S}_2, \tilde{\psi}\right)$, and for any $\mathbf{a} \in \mathfrak{S}_1$ and $\mathbf{b} \in \mathfrak{S}_2$, $\arg\min_{\mathbf{y}\in\mathfrak{S}_2} \tilde{\psi}\mathbf{ay}$ and $\arg\min_{\mathbf{x}\in\mathfrak{S}_1} \tilde{\psi}\mathbf{xb}$ are well-defined functions, and

$$\arg\min_{\mathbf{y}\in\mathfrak{S}_2} \tilde{\psi}\mathbf{ay} = \mathbf{b} \Longleftrightarrow \arg\min_{\mathbf{x}\in\mathfrak{S}_1} \tilde{\psi}\mathbf{xb} = \mathbf{a}. \tag{5}$$

[$\arg\min_x f(..., x, ...)$ is the value of x at which f reaches its global minimum.] Stimuli $\mathbf{a}$, $\mathbf{b}$ related to each other by (5) are called each other's *Points of Subjective Equality*, PSE. This law was first formulated in Dzhafarov (2002d) and was later shown to have unexpectedly strong consequences for a variety of issues, including the (im)possibility of modeling same-different discriminations by random images selectively attributable to stimuli being compared (Dzhafarov, 2003a-c). We view the Law of Regular Minimality as the most fundamental law of sensory discrimination (for empirical evidence and detailed discussion, see Dzhafarov 2002d; Dzhafarov & Colonius, 2005a, in press a).

Due to Regular Minimality the stimuli in $\mathfrak{S}_1, \mathfrak{S}_2$ can always be relabeled so that any two mutual PSEs are assigned identical labels (for details, see Dzhafarov & Colonius, 2005a). Denoting the set of these common labels by $\mathfrak{S}$, we transform $\left(\mathfrak{S}_1, \mathfrak{S}_2, \tilde{\psi}\right)$ into a *canonical discrimination system* $(\mathfrak{S}, \psi)$, in which Regular Minimality holds in its simplest, canonical form:

$$\arg\min_{\mathbf{y}\in\mathfrak{S}} \psi\mathbf{ay} = \mathbf{a} \text{ and } \arg\min_{\mathbf{x}\in\mathfrak{S}} \psi\mathbf{xb} = \mathbf{b}. \tag{6}$$

Dealing with a canonical system $(\mathfrak{S}, \psi)$ one should keep in mind that the physical identity of a stimulus is encoded by its label and its observation area (ordinal position within a pair); in particular, the two $\mathbf{a}$'s in $(\mathbf{a}, \mathbf{a})$ may be physically different.

Due to (6) the *canonical psychometric increments* $\Psi^{(1)}\mathbf{ab} = \psi\mathbf{ab} - \psi\mathbf{aa}$ and $\Psi^{(2)}\mathbf{ab} = \psi\mathbf{ba} - \psi\mathbf{aa}$ satisfy (1), i.e., they are deviation functions. We obtain the present ("universal") version of Fechnerian Scaling by assuming, in addition, that $\Psi^{(1)}$ and $\Psi^{(2)}$ are dissimilarity functions, and defining the Fechnerian metric in $\mathfrak{S}$ as the metric G induced by either of them in accordance with (4). It is irrelevant which of the two, $\Psi^{(1)}$ or $\Psi^{(2)}$, is used, they induce one and same metric G (in the previous versions of our theory this proposition was called the Second Main Theorem of Fechnerian Scaling). In terms of the (reduced) discrimination system $\left(\mathfrak{S}_1, \mathfrak{S}_2, \tilde{\psi}\right)$ this means that if $\mathbf{a}', \mathbf{b}' \in \mathfrak{S}_2$ are the PSEs for, respectively, $\mathbf{a}, \mathbf{b} \in \mathfrak{S}_1$, then the subjective distance between $\mathbf{a}$ and $\mathbf{b}$ in $\mathfrak{S}_1$ is the same as the subjective distance between $\mathbf{a}'$ and $\mathbf{b}'$ in $\mathfrak{S}_2$. The topologies based on open $\Psi^{(1)}$-balls and open $\Psi^{(2)}$-balls coincide, and both ψ and G are (uniformly) continuous in this topology.

Fechnerian Scaling with Transformed Probabilities

One of the open problems in our theory (or at least in the interpretation of Fechnerian distances as subjective distances) is that all Fechnerian computations applicable to $(\mathfrak{S}_1^*, \mathfrak{S}_2^*, \psi^*)$ are equally applicable to $(\mathfrak{S}_1^*, \mathfrak{S}_2^*, T\psi^*)$, where $T\psi^*$ is any monotone continuous transformation of ψ^*. For a discussion of how this problem can be tackled by, e.g., relating it to response bias models, see Dzhafarov & Colonius, 2005a, in press b). Here we only mention this issue to point out that the psychometric increments derived from $T\psi^*$ may very well fall outside $[0,1]$ or another finite interval of reals, because of which the notion of a dissimilarity function has to be generalized from $D : \mathfrak{S} \times \mathfrak{S} \to [0,1]$ to $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}^+$.

The generalization is as follows. A function $D : \mathfrak{S} \times \mathfrak{S} \to \mathbb{R}^+$ is a deviation function if it satisfies (1). It is a dissimilarity function if, in addition,

1. it satisfies (2) for any bounded sequences $\{\mathbf{a}_n\}_{n\in\mathbb{N}}$ and $\{\mathbf{b}_n\}_{n\in\mathbb{N}}$
   (a sequence $\{\mathbf{a}_n\}_{n\in\mathbb{N}}$ is bounded if, for some $\mathbf{a}$ and M, $D\mathbf{a}\mathbf{a}_n \leq M$);
2. for any sequence of chains $\left\{\mathbf{a}\mathbf{x}_1^n...\mathbf{x}_{k_n}^n\right\}_{n\in\mathbb{N}}$ it satisfies (3) and

$$D\mathbf{a}\mathbf{x}_{k_n}^n \to \infty \Longrightarrow D\mathbf{a}\mathbf{x}_1^n...\mathbf{x}_{k_n}^n \to \infty. \tag{7}$$

The theory essentially remains the same, except that the convergence and uniform continuity statements now generally have to be related to appropriate boundedness constraints.

Acknowledgement. This research was supported by NSF grant SES 0318010 to Purdue University.

References

Dzhafarov, E.N., & Colonius, H. (1999a). Fechnerian metrics in unidimensional and multidimensional stimulus spaces. Psychonomic Bulletin and Review, 6, 239-268.

Dzhafarov, E.N., & Colonius, H. (1999b). Fechnerian metrics. In P.R. Kileen & W.R. Uttal (Eds.), Looking Back: The End of the 20th Century Psychophysics (pp. 111-116). Tempe, AZ: Arizona University Press.

Dzhafarov, E.N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: Basics. Journal of Mathematical Psychology, 45, 670-719.

Dzhafarov E.N. (2001). Fechnerian Psychophysics. In N.J. Smelser, P.B. Baltes (Eds.) International Encyclopedia of the Social and Behavioral Sciences, v. 8 (pp. 5437-5440). New York: Pergamon Press.

Dzhafarov, E.N. (2002a). Multidimensional Fechnerian scaling: Regular variation version. Journal of Mathematical Psychology, 46, 226-244.

Dzhafarov, E.N. (2002b). Multidimensional Fechnerian scaling: Probability-distance hypothesis. Journal of Mathematical Psychology, 46, 352-374.

Dzhafarov, E.N. (2002c). Multidimensional Fechnerian scaling: Perceptual separability. Journal of Mathematical Psychology, 46, 564-582.

Dzhafarov, E.N. (2002d). Multidimensional Fechnerian scaling: Pairwise comparisons, regular minimality, and nonconstant self-similarity. Journal of Mathematical Psychology, 46, 583-608.

Dzhafarov, E.N. (2003a). Selective influence through conditional independence. Psychometrika, 68, 7-26.

Dzhafarov, E.N. (2003b). Thurstonian-type representations for "same-different" discriminations: Deterministic decisions and independent images. Journal of Mathematical Psychology, 47, 208-228.

Dzhafarov, E.N. (2003c). Thurstonian-type representations for "same-different" discriminations: Probabilistic decisions and interdependent images. Journal of Mathematical Psychology, 47, 229-243.

Dzhafarov, E.N. (2004). Perceptual separability of stimulus dimensions: A Fechnerian approach. In C. Kaernbach, E. Schröger, H. Müller (Eds.), Psychophysics beyond Sensation: Laws and Invariants of Human Cognition (pp. 9-26). Mahwah, NJ: Erlbaum.

Dzhafarov, E.N., & Colonius, H. (2005). Psychophysics without physics: A purely psychological theory of Fechnerian Scaling in continuous stimulus spaces. Journal of Mathematical Psychology, 49, 1-50.

Dzhafarov, E.N., & Colonius, H. (2005). Psychophysics without physics: Extension of Fechnerian Scaling from continuous to discrete and discrete-continuous stimulus spaces. Journal of Mathematical Psychology, 49, 125-141.

Dzhafarov, E.N., & Colonius, H. (in press, a). Reconstructing distances among objects from their discriminability. Psychometrika.

Dzhafarov, E.N., & Colonius, H. (in press, b). Regular Minimality: A fundamental law of discrimination. In H. Colonius & E.N. Dzhafarov (Eds.), Measurement and Representation of Sensations. Mahwah, NJ: Erlbaum.

Dzhafarov, E.N., & Colonius, H. (in press, c). Generalized Fechnerian Scaling. In H. Colonius & E.N. Dzhafarov (Eds.), Measurement and Representation of Sensations. Mahwah, NJ: Erlbaum.

SYSTEMS FACTORIAL TECHNOLOGY ANALYSIS OF POMERANTZ'S CONFIGURAL FIGURES

Ami Eidels, James T. Townsend, Indiana University
James R. Pomerantz, Rice University
e-mail: aeidels@indiana.edu

Abstract

People are especially efficient in processing certain visual stimuli such as human faces or 'good configurations'. These stimuli are presumably processed holistically (or configurally): we process not only their components, but also spatial relations among these components. Novel properties such as closure or symmetry may emerge when we combine the features into a unified configuration. We attempted to define the properties of the cognitive system that underlies the processing of configural figures, by employing Townsend and Nozawa's (1995) systems factorial technology. Our data emphasize the importance of context-items and their topological similarity to the target-item (even when the context-items and the target are not concurrently displayed).

People detect faster, and more accurately, a certain feature when it is presented within a context, then when presented alone (Pomerantz, Sager, & Stoever, 1977). Similarly, people seem to better identify a given feature, say a nose, within the context of a face, than alone (Tanaka and Farah, 1993). What is so special about these contexts? And, more generally, what is the special thing about faces and "good figures" that aids in processing their constituting features? According to one view, highly symmetrical ("good figures") or highly learned (faces) figures are processed as a Gestalt. That is, they are processed as a whole, and it is that holistic experience that aids in the resolution of their constituent parts. Another view suggests that combining features in a certain way involves the creation of emergent features, such as symmetry or closure. But how exactly does a Gestalt, or an emergent feature, operate to facilitate performance in the examples that we have described?

Pomerantz and his colleagues (1977) provided an appealing demonstration of the salutary effect of context. Discriminating between orientations of diagonal lines with no context added (Figure 1 – Panel A) took participants 1884 ms. The addition of a context (Panel B) that provided no task relevant information, but contributed to closure or symmetry, reduced response times in the composite condition (Panel C) to as little as 749 ms.

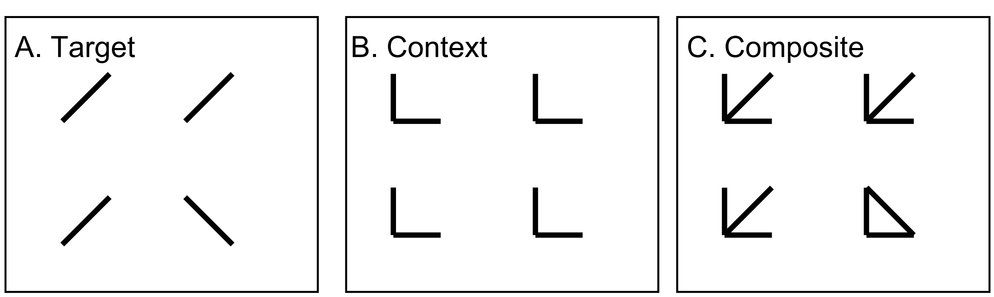


Figure 1: Stimuli used by Pomerantz et al (1977) to produce configural superiority effect. Localization of the odd item was faster in display C then A (749 vs. 1884 ms, respectively).

Given Pomerantz's figures, the features seem to somehow integrate into a unified whole, a Gestalt, which in turn makes the processing of its nesting features more efficient. What is the nature of that Gestalt and can it be measured using the existing analytical tools? Townsend and Nozawa (1995; see also Wenger and Townsend, 2000) developed a methodology, augmented by a related mathematical theory, that enables us to characterize the cognitive system in terms of architecture (*serial* vs. *parallel* vs. *co-active*), capacity (*limited* vs. *unlimited* vs. *super*), stopping rule (*minimum time* vs. *exhaustive*) and (in)dependency. In the current study we applied System Factorial Technology (SFT) in the attempt to decipher how configurality, created by the spatial arrangement of features, affects processing. Specifically, we examined if the influence of configuration can be measured in terms of capacity, architecture, and stopping rule.

To apply Townsend and Nozawa's SFT, one should "tame" the stimuli of interest into a redundant target search task. As stimuli, we employed figures, constructed of angles and oblique lines, which were shown by Pomerantz et al to produce a noticeable configural superiority effect. Each participant performed in a redundant target (OR) detection task, where she or he had to detect the presence of certain features that were designated earlier as targets. If the constituent features co-act (or interact) to create a unified whole, then presenting two targets together should produce a marked redundant-target effect. This facilitation of RTs in response to displays containing configuration should be bigger than would be expected on the basis of summing the effects of distinct components. A plausible explanation for such results could then be formulated in terms of super capacity and co-activation. In that case, one should expect violations of Miller's inequality, as well as $C(t)$ values higher than 1 [see Townsend & Nozawa (1995) for a thorough explication of $C(t)$ and other SFT measures].

We expect topological similarity to play a major role as well. Topology deals with properties of geometric figures that are not changed by homeomorphism, such as stretching or bending. Any two triangles are topologically equivalent. Topologically similar items would suffer relatively slow RTs if they are assigned to different responses, since discrimination between them would be difficult and requires additional processing time. On the other hand, topologically similar items that are assigned to the same response (especially if they are topologically different from the other items that may be presented) would be processed quickly, as little discrimination is needed.

Method

The stimulus set was based on the figures employed by Pomerantz, Sager and Stoever (1977). We used the factorial combination of a diagonal line (either left, "\", or right, "/") and a right angle (open either to the right, "L", or to the left). The stimuli were presented in black over a gray background. The viewing distance was approximately 50 cm from the center of the screen, such that stimuli subtended a visual angle of 1.5 degrees. On each trial, a fixation cross appeared in the center of the screen for 500 ms., followed by a blank screen (also 500 ms.). Then, a single stimulus was presented in one of four possible locations around the center (this trial-to-trial spatial uncertainty prevents adaptation and attention to a particular location), until a response was made. Participants responded by pressing one of two mouse keys pre-designated as YES or NO.

We conducted several experiments in which we varied the spatial relations between the two features thus manipulating the properties of closure or proximity, and also the topological (dis)similarity among the items. Stimuli sets from experiments 1-3 are illustrated in Figure 2. In all experiments, "/" and the mirror image of "L" were defined as targets. Therefore, when either one, or both, appeared (Figure 2 – a, b, & c) – the participant had to respond YES. If only distractors appeared (Figure 2 - d) – then the correct response was NO. We had 4-8 participants in each experiment; each performed on 1280 experimental trials (320 per condition).

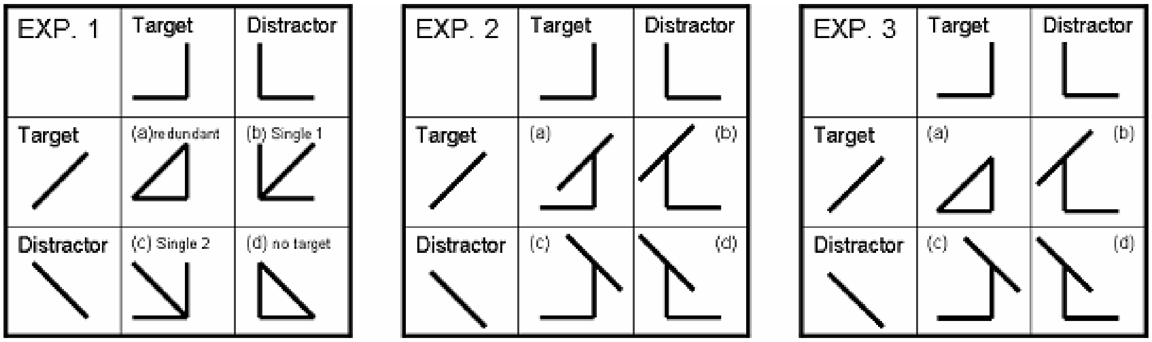


Figure 2: The stimuli-sets for experiments 1, 2, & 3. Factorial combinations of target and distractor yield redundant- (a), single-target 1 (b), single-target 2 (c) and no-target (d) conditions.

In Experiment 1, the two single-target stimuli were constructed such that they shared the same topology and were assigned to the same response (Figure 2, left panel - b & c). The redundant- and no-target stimuli are topologically similar to each other (a & d), but are assigned to different responses. Also, they can be considered 'good' configurations, as they both possess the property of closure. In Experiment 2, we created the stimuli such that they would all belong to the same topological class. Note that none of the items is more (or less) closed or symmetrical than the others. In Experiment 3, three of the figures remained topologically similar to each other, whereas the features of the redundant-target stimulus were spatially arranged to create a closed form (a triangle).

Results and Discussion

In all of the experiments, error rate was low for all participants and across conditions and no RT-accuracy trade off was observed. Analysis of data was restricted to correct responses only. In Experiment 1, responses on single target trials appeared to be the fastest (448.6 and 451.1 ms for single-target 1 and 2, respectively). The no-target condition was slower, with 598.9 ms, and the redundant target condition was the slowest, averaging 634.1 ms across all participants. Further examination of the data reveals that exactly the same ordering was maintained for each of the individual performers. The factors determining the efficiency of processing seem to be the <u>topological (dis)similarity</u> among items, and the response allocation: the two triangles that constitute the redundant- and the no-target displays belong to the same topological class, yet call for different responses. Discrimination between the two 'closed' figures had to be made, causing in turn slower responses. A reversed redundant-target effect is observed. Hence, the 'good' property of closure does not necessarily facilitate processing.

In Experiment 2, responses on redundant target trials appear to be the fastest (558.5 ms). The two single-target conditions were slower (654.2 and 591.5 ms) and the no-target condition appears to be slowest, averaging 696.8 ms across all participants. Since all stimuli were carefully constructed as members of the same topological class, no stimulus was more 'configural' than any other; i.e., all four types of stimuli had more or less the same spatial relations between the oblique line and the angle. Under these conditions, the usual ordering of RTs was obtained. No violation of Miller's inequality was observed and only one participant (P 3) violated the lower bound set by Grice inequality. Capacity, gauged by the capacity coefficient, C(t), was between 1.0 to .5; these values indicate that the system had an unlimited to moderately limited capacity.

In Experiment 3, the mean RT on the redundant-target trials was the fastest (410 ms). The two single-target conditions were slower (508 and 478 ms) and the no-target condition was the slowest (606 ms). The same ordering was kept for each of the individual performers. Responses on the topologically unique redundant-target displays were much faster than on the single target displays. Indeed, C(*t*) values go way above 1.0, suggesting super-capacity.

In Experiment 1, 2, & 3, we changed the relative position of the line and angle such that some combinations would yield closed figures and some would yield unclosed lines. As a result, some figures had holes and were therefore topologically different from other figures that did not have holes. By manipulating the topological (dis)similarity among display conditions we were able to obtain, eliminate, and reverse the RTE. Our results suggest that it is not the configuration of each item per-se, but rather its similarity or dissimilarity to the other set-items that governs the relative speed of processing.

Previous studies have already shown that the resemblance between the target item and other items (distractors) on the same display determined whether or not the target would 'pop-out'. In the current design, rather than simultaneously presenting the target- and distractor-items, only a single item was presented per trial. Nonetheless, the non-displayed items ('implicit-context'), i.e., other items from the stimuli-set that appeared on previous or subsequent trials, were just as influential in determining the target-detection latencies. Our findings are consistent with a principle notion of information theory (e.g., Garner, 1974), arguing that perception depends not only on the stimulus presented for view but also on its alternatives, those stimuli that could have been presented although were not presented on that particular trial.

Acknowledgements

This work was supported NI-NIMH grant MH 057717. We would like to thank Leslie Blaha, Mario Fific, and Chris Honey for their comments; and, Christi Gibson and Matt Primeau for their assistance in collecting and analyzing the data.

References

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422-435.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology: General. 130,* 534-543.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39,* 321-359.

Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *The Journal of General Psychology, 127*, 67-99.

SUBJECTIVE MAGNITUDE AS A FUNCTION OF NUMERICAL MAGNITUDE AND NOTATION: ADDITIVE EFFECTS?

Daniel Fitousi and Daniel Algom
Department of Psychology,
Ramat Aviv, Tel-Aviv University, Israel
dannyfit@post.tau.ac.il, algomd@freud.tau.ac.il

Abstract

We applied additive factors logic (Sternberg, 1969) to study the effects of order of magnitude and notation on the perception of numerical magnitude. Participants judged the magnitude of single-, double-, and three-digit numbers appearing either in Arabic or word notation. Numbers appearing in Arabic notation were responded faster than those presented in words. The times to judge the size of Arabic digits increased for multi-digit compared with single-digit numbers, but were approximately the same for two- and three-digits. For number words, judgment time increased in a linear fashion with order of magnitude.

Consider the string of letters, "FOUR." Is it a word? Is it a number? It is obviously both. If it is a number, is it processed differently than "4?" The answer to this simple question seems to depend on the task at hand. If people are asked to simply read "FOUR" and "4", then they are faster with the former than with the latter. However, when people are asked to judge the magnitude, they respond faster to "4" than to "FOUR." This effect of notation has been found with single digits. Does the same pattern hold with two- and three-digits numbers?

Research on numerical cognition has been notoriously uneven, with a disproportionate majority of the pertinent studies concerned with single-digit numbers. Perception of multiple-digit numbers might be different, involving syntactic, lexical and combinatorial operations with the composite digits. Such processes are gratuitous with single-digit numbers. Consequently, single- and multiple-digit numbers might be processed differently. One of the aims of the current research was to probe the effect of notation with single-, two-, and three-digit numbers.

There exists indeed evidence suggesting that single- and multiple-digit numbers are not processed by the same mechanism. When people select the numerically larger member of a pair of digits, the larger the numerical difference, the shorter is the RT. This distance effect (Moyer and Landauer, 1967) is invariably found with single digits but often not found when people compare two-digit numbers (Brysbaert, 1995; Verguts & De Moor, 2005). And, the effect seems to vanish for three-digit numbers (Hinrichs, Berie, & Mosell, 1982; Poltrock & Schwartz, 1984).

The different cognitive operations with single- and multiple-digit numbers are also reflected in compatibility effects observed when people compare two-digit numbers (Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Nuerk, Weger, and Willmes, 2001). The pair 42-58 is compatible because 4<5 and 2<8, whereas the pair 29-63 is incompatible because 2<6 but 9>3. Decision times decrease and accuracy improves for pairs in which the decades and the units are compatible even when numerical distance is held constant. Lexical and combinatorial operations are obviously missing from single-digit processing.

Consider notation, again. For naming, Fias, Reynvoet, and Brysbaert, (2001) found that an irrelevant number word interferes with the naming of an Arabic digit, although naming (reading) a number word did not suffer interference from an irrelevant Arabic digit. Therefore, in a naming task, the "word" properties of a number word may well overshadow its "number" properties. An Arabic digit, in contrast, may well function as a picture with a privileged access to its semantic meaning. If so, this asymmetry in representation only holds for single digits. It is highly unlikely that two- and three-digit numbers have a unique picture-like representation.

However, when the task is that of parity judgment (deciding whether a number is odd or even) or of magnitude (whether or not the digit is larger than 5), then the responses are faster for Arabic digits than for number words (Damian, 2004; Fias, et al, 2001). According to an influential model by Dehaene (1992), people must convert number words to Arabic digits to extract parity and judge magnitude in an efficient fashion.

In the present study we addressed some of these questions by applying Sternberg's (1969) additive factor logic with notation and order of magnitude as factors. In one block, the numbers between 1and 999 appeared as words. In another block, the same numbers were presented as Arabic digits. On each trial the participant decided whether the presented number was larger or smaller than a standard. Single-digits were compared to 5, two-digit numbers were compared to 50, and three-digit numbers were compared to 500. We asked: Are judgments of magnitude to Arabic digits faster than those made to the same numbers presented as words? If so, is the advantage persevered across order of magnitude? Does order of magnitude make a difference? If RT is longer, does it increase at the same rate across notation?

Method

<u>Population</u>: Eight young volunteers from the Tel-Aviv university community participated.
<u>Stimuli and Apparatus:</u> We selected 9 numbers from each of the following ranges: 0-9, 10-99, 100-999 in a random fashion. All units, decades and hundreds in the 1-999 range were represented, resulting in 27 numbers. In one block of trials, the numbers appeared in an Arabic notation (font David, size 24). In another block, the same numbers appeared in word notation (font David, size 24). The stimuli were generated by an IBM compatible microcomputer (PC 486) and displayed on a regular screen. The numbers appeared black over a white background at the center of the screen. To avoid adaptation, we introduced a trial-to-trial spatial uncertainty of up to 15 pixels around the target location. The viewing distance was 50 cm.
<u>Procedure</u>: The participants were tested in a dimly lit room. Each participant performed in a block with number words and in a block with Arabic digits. Half of the participants first performed in the former block and half first performed in the latter block. Participants were instructed to indicate whether the presented number was greater or smaller than the appropriate standard by pressing either a right- or a left-hand key on the computer keyboard. Single-, two-, and three digit numbers were compared with the standards 5, 50, 500, respectively. Key assignments were balanced across blocks and participants. Each number in each block was presented twice, resulting in a total of 54 trials per block. Twenty practice trials unbeknownst to the participant preceded the experimental trials in each block. The stimuli were response terminated. A new number was presented following a 500 ms interval. Trials in which RT was either greater than 1500 or smaller than 150 ms were excluded from the analyses.

Results and Discussion

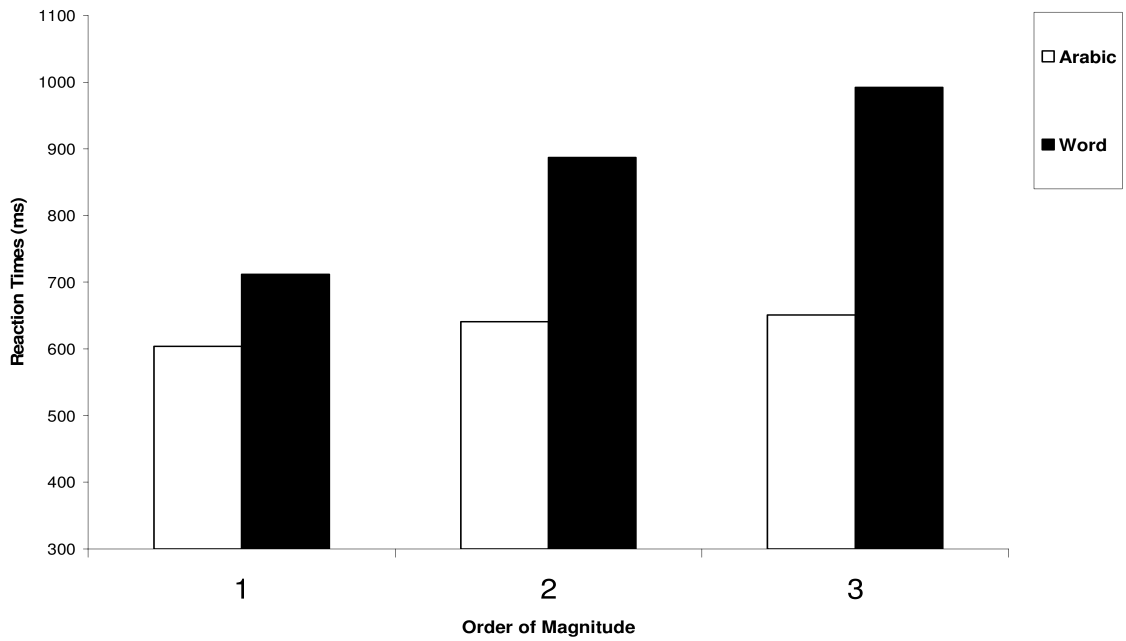


Fig.1. Mean reaction times to decide numerical magnitude for different notations and orders of magnitude.

Figure 1 depicts the effects of notation and order of magnitude on the time to decide whether the number was larger or smaller than the standard. Salient is the advantage of Arabic over word notation across all stimuli. Arabic digits were processed 136 ms faster than their corresponding words [F (1, 7) =6.78, MSE=99246, $p<0.05$]. These results extend previous observations for single-digits (Damian, 2004). Consider next order of magnitude. The higher the order of magnitude of a number, the longer it takes to judge its magnitude [F (1, 7) =6.44, MSE=18432, $p<0.05$]. Each additional digit or word takes a toll on performance (see also Poltrock & Schwartz, 1984). It is also apparent in Figure 1 that the advantage of Arabic over word notation does not remain constant, but increases with order of magnitude. Arabic numbers themselves do not exhibit a dramatic increase across order of magnitude, whereas number words do. Double- and three-digit Arabic numbers are processed approximately at the same speed, although they are slower than single-digit numbers. For words, judgments increase linearly with magnitude. These trends are documented by the notation x order of magnitude interaction [F (2, 14) =5.98, MSE=9368, $p<0.05$].

Discussion

Why are Arabic digits judged faster than words? According to Dehaene's (1992) influential model, Arabic digits are dedicated human's means for processing magnitude. Therefore, the advantage of Arabic digits may reflect the time it takes to convert number words to Arabic digits. Why does magnitude affect Arabic digit processing? The effect is small and might derive from what is generally known as problem size effect (Ashcraft, 1992). The larger the number, the longer are any arithmetic operations concerning that number. In addition, single digits may be processed especially fast due to frequency, and due to unique picture-like representation. Why does magnitude affect judgments of magnitude of number words? One reason is trivial: Larger numbers are composed of a greater number of words. In addition, the conversion to Arabic form takes proportionally longer to accomplish with larger numbers. These various effects conspire to produce the present main effects as well as their interaction in processing numbers in the two notations.

References

Ashcraft, M.H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition*, *44*, 75-106.

Brysbaert, M. (1995).Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. *Journal of Experimental Psychology: Human Perception & Performance*, *2*, 435-447.

Damian, M. (2004).Asymmetries in the processing of Arabic digits and number words. *Memory & Cognition*, *31*, 164-171.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*. *44*. 1-42.

Fias, W., Lammertyn, J., Reynvoet, B., Duponnt, P., & Orban, G.A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, *15*, 47-56.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task? *Psychological Research*, *65*, 242-249.

Hinrichs, J.V., Berie, J.L., & Mosell, M.K. (1982). Place information in multidigit number comparison. *Memory & Cognition*, *10*, 487-495.

Moyer, R.S., & Landauer, T.K. (1967). Time required for judgments of numerical inequality. *Nature*, *215*, 1519-1520.

Nuerk, H.C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, *82*, B25-B33.

Poltrock, S.E., & Schwartz, D.R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 32-45.

Sternberg, S., (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315.

Verguts, T., & De Moor, W. (2005). Two-digit comparison: Decomposed, holistic, or hybrid? *Experimental Psychology*, *52*, 1-6.

COMPARING ODDBALL AND FREE CONTEXT ERP PARADIGMS IN EVALUATIVE TASKS

Isabel B. Fonseca[1], Armando M. Oliveira[2], Marta Teixeira[2], Eduardo Santos[2], Fátima Simões[3]
[1]FPCE-University of Lisbon, Portugal; [2]Institute of Cognitive Psychology – University of Coimbra, Portugal; [3]University of Beira Interior, Portugal
isabelbf@fpce.ul.pt ; l.dinis@fpce.uc.pt

Abstract

This study contrasts the results obtained with a modified oddball paradigm and a random, free context paradigm, in an affective picture evaluation task organized as a 3 valence x 2 arousal factorial design. 15 pictures selected from IAPS were used in the oddball paradigm to create a "positive low-arousal" context, while 18 others were employed as embedded targets. These same targets were also employed in a random, free-context paradigm. Same specifications were used regarding EEG derivations, physiological data collecting, stimuli presentation and instructions. Consistent with a "liberation from context" expectation, comparative increases of peak amplitude in positive conditions were observed with the random paradigm, accompanied by decreased peak latencies. Higher overall amplitudes were found with the oddball task, suggesting that effects of "evaluative distance" and "intrinsic evaluation" combine. The peculiar finding of lowered amplitude in the low arousal negative condition was shown robust across paradigms.

Studies revealing increased late cortical positive activity (LPP) in affective categorization tasks have relied mostly upon variants of the classical oddball procedure (Cacioppo *et al.,* 1994; Crites *et al*., 1995). One line of debate was soon opened regarding the modulating factor of the LPP activation: positive-negative valence, to which researchers like Cacciopo incline to, or arousal, as empirically championed by Cuthbert (Cuthbert et al, 2000). In an attempt to disentangle each factor's effect, at the same time allowing for their possible joint contribution, we conducted a previous 3 valence (positive, neutral, negative) x 2 arousal (high, low) factorial experiment. Results suggested that both factors had an impact on the ERP component. However, having obeyed an oddball paradigm (with a positive low-arousal context), these results were vulnerable to two related critics: a confounding between evaluative distance and intrinsic valence/arousal, and an oversimplifying view of LPP multiple functional roles and neurological subbasements (Goldstein *et al*., 2002).

To address these issues, the same two-factor experiment was replicated with a free-context ERP paradigm devised by Schupp *et al.* (2000), resting upon random presentation of equal-probability classes of affective pictures. Valid comparisons between results arising from different paradigms are often hindered by the use of different stimuli, procedures, or designs. All these factors were kept constant in our two experiments, with the exception of specific procedural aspects required by each paradigm. This offers the double chance of identifying specific paradigm-dependent effects, of potential methodological and substantive significance, as well as common invariant findings, whose robustness can thus be convincingly established.

Method

Subjects. 12 undergraduate students (4M, 8F) at the University of Coimbra participated in the oddball experiment. A second group of 10 students (3M, 7F) volunteered to the free-context task.

Stimuli. **I.** Oddball task**.** 15 pictures selected from IAPS (1999) on the basis of normative ratings provided an overall "positive-low arousal" context; 18 others were used as embodied targets, 3 in each valence x arousal condition. **II.** Free-context task: the same 18 target pictures added with further 18, all embodying a similar 3 valence x 2 arousal factorial structure, resulting in 6 equal-probability classes of affective stimuli.

Common EEG assemblage and analysis. Three monopolar EEG derivations referenced to the right mastoid were placed at sites Fz, Cz and Pz (10-20 IS), supplemented with an EOG channel for artifacts control. Sample rate was of 200 Hz in both tasks, and the same band-pass filter of 0.1-35 Hz was used. Time epochs included in both cases a 150 ms pre-trigger period, extending for 1 s after picture onset. EEG sweeps were separately averaged for each of the six experimental conditions at each recording site. After baseline correction to the pre-trigger period, waveforms were low-pass filtered at 10 hertz. Maximum positive peak amplitude and peak latency were measured in the 300-700 ms window following stimuli onset.

Procedure. Participants sat in a recliner, in dim light, in front of a VGA monitor (50 com ahead). The required response (evaluation of pictures as positive, neutral, or negative, through pressing a button after stimulus offset) as well as general instructions were invariant for all subjects in both tasks. **I.** Oddball task: 360 series of five slides were presented, each composed of four "context" pictures and one target. The first two pictures in the series were always "positive-low arousal", with targets occurring equally often at each remaining position. Pictures were displayed for 1 second each (ISI:1,5 s). **II.** Free context task: the 6 pictures pertaining to each experimental condition were randomly presented across 1080 trials. The production of local contexts was avoided by using partially random series organized in blocks, which were on their turn randomized. Regarding all other aspects, procedures were the same as in **I**.

Data analysis. Data were analysed through separate repeated-measures ANOVAs performed on peak amplitude (microvolts) and peak latencies (ms) for each EEG derivation (within-subject factors: valence and arousal). However, only amplitude results will be referred hereafter.

Results

Fig. 1 plots "free-context" amplitude results, on the left, and "oddball" results on the right, organized by EEG derivation (Pz, Fz, Cz, from top to bottom). The most salient findings are the comparatively heightened amplitude of the positive valence condition on the free-context task (mostly at Pz), and the sensibly decreased amplitude of the negative high-arousal condition that accompanies it. These findings are best expressed in Figure 2 by the crossovers around the neutral stance, and have different meanings. The first one indicates an inhibition of positive modulation of the LPP due to "context congruity"; the second one, that a considerable portion of LPP activation by high-negative valence was in fact due to "context incongruity".

A second observation is the substantially diminished ERP amplitude displayed for the negative-low condition across all graphics in both tasks. This finding, never to our knowledge reported (despite undeniably robust), probably owns its expression to the particular demands of the design regarding picture's contents sampled in IAPS. Studies asking for discrete emotions classification of IAPS's pictures have revealed a predominance of "sadness" on the negative low aroused quadrant. It cannot be excluded that LPP activation may, in some circumstances, be responding to discrete emotion categories rather then to dimensional factors such as valence and arousal.

As in the former oddball experience, suggestions for a differential modulation of ERP amplitude as a function of scalp location are found in the free-context task. While arousal effects (and valence*arousal interaction) appear as statistically significant at Pz and Cz, valence seems to express preferentially at Fz (see results for ANOVA on the middle left panel). This concurs with extended evidence available regarding valence-related effects on frontal electrophysiological activity.

Arousal differences observed for positive and neutral valence on the oddball paradigm seem basically dependent on "distance effects", and vanish in the free-context task. Also, for these valence categories, increased "distance to context" in arousal was signalled by amplitude depletion.

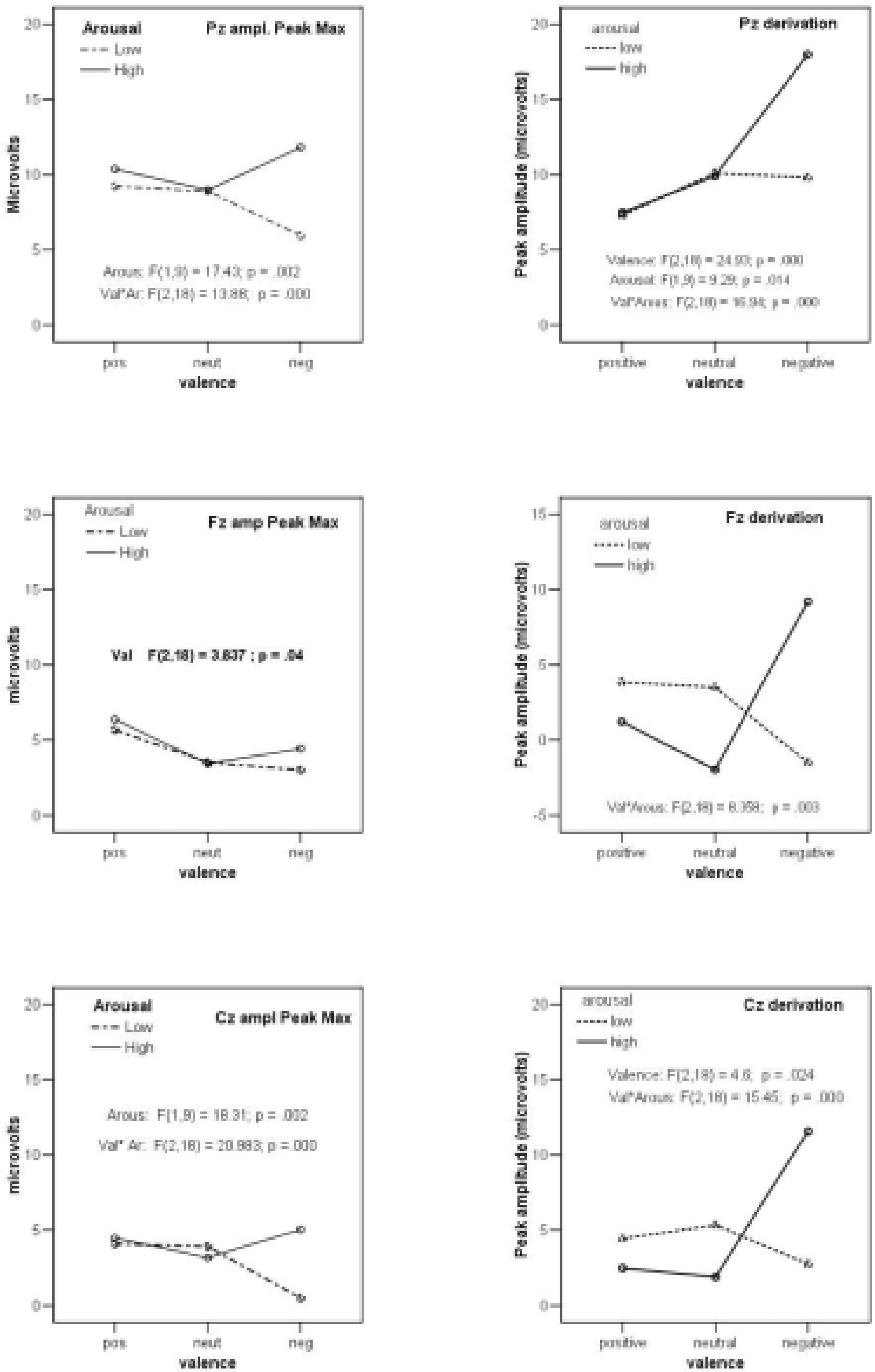


Figure 1: Factorial plots for peak amplitudes at Pz, Fz, Cz

Left panel: free context task **Right panel:** oddball task

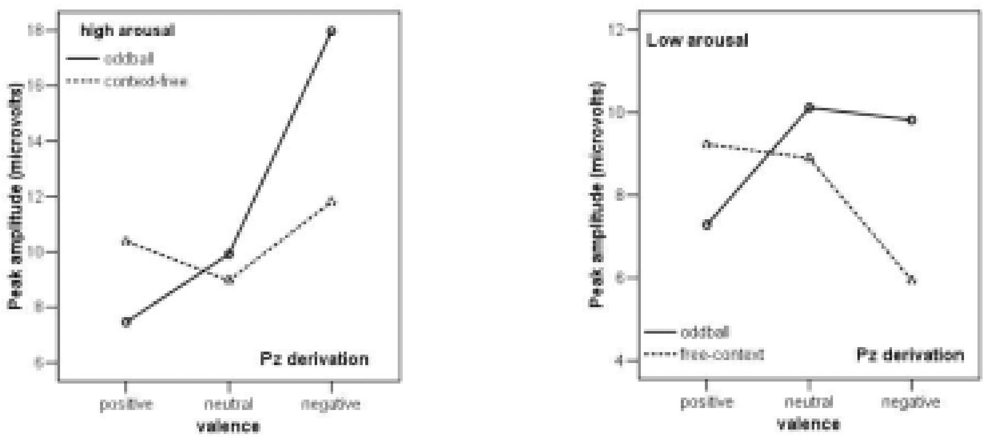


Fig 2. Comparison between oddball and free-context plots at Pz derivation

Left panel: high arousal **Right panel:** low arousal

Discussion

As expected, liberation from "reduced distance to context" was apparent in the positive conditions. This is most clearly illustrated by the crossovers displayed in Fig. 2 (Pz site). As for arousal, "increased distance to context" of the oddball seems to have translated in lowered amplitudes for the positive and neutral conditions. This is supported by the near constancy of low arousal amplitudes for both valence categories across tasks, together with the almost complete overlap of their high and low arousal levels displayed on the free-context task. LPP activation by arousal might thus require a "context effect" to be present whenever positive/neutral evaluations are at stake.

The much larger amplitudes observed for the negative-high arousal conditions in the "oddball" task entail a suggestion that "distance effects" and "intrinsic affect" add to each other. This is compatible with the increasingly recognized complex functional role of the "late positive complex.

The peculiar finding of marked amplitude reduction for negative-low arousal pictures was proven reliable across tasks. While it remains an open issue, two interconnected interpretations are that it may be a specific signature for "sadness" (a discrete emotion) and/or an expression of a more general protecting mobilization-minimization mechanism (Taylor, 1991).

References

Cacioppo, J., Crites, S., Gardner, W., and Berntson, G. (1994). Bioelectrical echoes from evaluative categorization: I. A late positive potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology, 67(1),* 115-125.

Cuthbert, B., Schupp, H., Bradley, M., Birbaumer, N., and Lang, P. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological Psychology, 52,* 95-111.

Goldstein, A., Spencer, K., Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and Novelty P3. Psychophysiology, *39*, 781-790.

Schupp, H., Cuthbert, B., Bradley, M., Cacioppo, J., Ito, & Lang, P. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. Psychophysiology, *37*, 257-261.

Aknowledgment. This research was supported by grant POCTI/PSI/41235/2001 (FEDER/FCT).

OLFACTORY SENSITIVITY MEASURES CAN PREDICT COGNITIVE IMPAIRMENT: A PARAMETRIC AND NONPARAMETRIC APPROACH.*

Ana Garriga-Trillo & Francisco Aguilera-Genicio
Centro de Envejecimiento y Enfermedades Neurodegenerativas-Investigación Psicofísica
(Centre for Ageing and Neurodegenerative Diseases-Psychophysical Research)
UNED, Madrid, Spain
agarriga@psi.uned.es ;http://www.uned.es/investigacion/institutos/CEEN/ana_julia/equipo_aj.html

Abstract

Deficits in olfactory sensitivity have been found in neurodegenerative diseases using various sensitivity measures. Chan et al. (2002), Garriga-Trillo (2003) and Susuki et al (2004) have used olfactory identification tests, adapted to their countries, as indicators for Alzheimer´s Disease. Not considering neurodegenerative diseases, this research will just study if olfactory sensitivity can predict cognitive impairment in an elderly Spanish sample (N=65) using both parametric and nonparametric techniques. Cognitive impairment was measured by the Mini-Mental State Examination (Folstein et al, 1975) and the Eurotest (Carnero, 2005). Olfactory sensitivity was measured by the number of correct odour identifications with a non-verbal task and by thresholds. Using a backward regression model and the AnswerTree classification system, our results show that: (1) Mini-Mental scores are predicted by the number of odours correctly identified, but (2) both olfactory thresholds and the number of odours correctly identified, can predict Eurotest´s scores. Therefore, olfactory sensitivity can predict cognitive impairment using both statistical analyses, being cognitive impairment predicted from both sensitivity measures when using the Eurotest.

When studying olfactory sensitivity measures within psychophysics, back in the 1980s, Cain (1981), Schiffman (1983) and Garriga-Trillo (1985, 1987) had mentioned that chemosensory defects, especially smell impairment, could be related to ageing and to serious conditions as neurological disorders. Olfactory tests were used by neurologists to diagnose brain damage. During that decade, greater attention was given to olfactory impairment and the smell clinics were established in the US with government support. In the 1990s, interest in olfactory dysfunction increased tremendously. Engen (1991) presents two approaches to the study of the senses. One considers classical sensory physiology and psychophysics and the other emphasizes the fact that peripheral information is related and controlled by higher levels in the nervous system. Although both approaches were considered to be in conflict at that time, he proposes a new approach linking both systems and exemplifies this issue considering the olfactory modality. An amplification of these ideas were proposed by Buck & Axel (1991) on odorant receptors and the organization of the olfactory system that leads to the conscious experience of a recognizable odour. Sensory and cognitive tasks in olfaction were linked. Further links to olfaction-cognition-ageing appeared in Moberg & Raz (1997), studying encoding strategies and cognitive abilities. The research also adds a connection, as in the 80´s, between neurological disorders or neurodegenerative diseases and olfactory functioning, Mesholam, Moberg, Mahr & Doty (1998) present a broad meta-analysis of olfactory functioning in neurodegenerative disease, considering in particular Alzheimer´s and Parkinson´s disease in English-language studies. The olfactory domains included in their analyses were odour identification, recognition and detection thresholds. The data were analyzed by classical non-parametric techniques (Mann-Whitney U test and Friedman´s two-way analysis of variance) with 43 studies. Results showed that both Alzheimer´s and Parkinson´s patients had severe olfactory deficits with all three olfactory measures compared with their control groups. However, no significance difference was found between Alzheimer´s and Parkinson´s patients in the three olfactory measures. New non-parametric techniques that create classification systems displayed in decision trees have not been used.

To rule out a cross-cultural effect in studies concerning olfactory sensitivity as indicators of Alzeimer´s disease, Chan et al (2002), Garriga-Trillo (2003) and Susuki et al (2004) have used olfactory identification tests adapted to their countries. Other cultural effects can also be induced by the cognitive impairment test used. Cantero et al (2002) and Cantero (2005) have developed a

Spanish (and also European) short, easy and ecological test to detect cognitive impairment and dementia based on the knowledge and handling of the Euro currency. Illiterate effects are then controlled.

Considering all the above mentioned studies, this research will try to add new perspectives in studying cognitive impairment (not neurodegenerative diseases) starting from olfactory sensitivity measures. Olfactory sensitivity will predict (not merely indicate) cognitive impairment. The predictions will be obtained both by the classical parametric method of a backward regression model and by AnswerTree creating decision trees from classification systems by Chi Square. For measuring olfactory sensitivity a classical measure (odour thresholds) and a new odour identification test controlling cultural and education effects will be used. For measuring cognitive impairment a classical test (MMSE) and the new Eurotest that controls cultural and education effects will be used.

Method

Participants

A total of sixty-five volunteer healthy elderly subjects living at home, in 10 different Spanish towns, took part in the experiment. All of them took the MMSE and only 26 took the Eurotest. Their **ages** ranged from 62 to 90 years (M = 75 years), and 45 of them were females. The distribution for **years of schooling** (YS) and **civil status** (CS) was: **YS -** no schooling at all (6.2%), elementary-six years (66.2%), high school-12 years (24.6%), university-more than 12 years (3.1%); and **CS -** single (16.9%), married (33.8%), widowed (44.6%), divorced (4.6%). All subjects were naive to the nature and aims of the experiments and none had any physical deficits linked to partial or total anosmia. All participants were tested individually in their city´s day centre for the elderly . They received no payment for participating in the experiment.

Materials and Stimuli

All subjects underwent two validated tests of cognitive impairment: the Mini-mental State Examination (Folstein et al, 1975) and the Eurotest (Carnero Pardo et al, 2002; Carnero, 2005). The Mini-mental State Examination (MMSE) used was a Spanish translation of Folstein´s et al original test. It is divided in two sections: the *first* requires vocal responses for orientation, memory and attention and the *second* concerns language (the ability to name and to follow verbal and written commands) and a copy of a complex polygon. Its overall score is 30 points. The highest the score obtained means less cognitive impairment. The Eurotest, as described by Carnero et al (2005) and Carnero (2005), is an easy, brief and useful test, based on the knowledge and handling of Euros coins. It can be applied to illiterate subjects and can detect cognitive impairment (Carnero & Montoro, 2004). It is divided in three parts: the first requires knowledge/denomination of the euros coins, the second requires calculation of different amounts of Euro coins and the third concerns a memory task of the previous coins shown. Its total score is 35 points. The highest the Eurotest score obtained means less cognitive impairment. Three other tests were used: a 20 item picture identification test (PIT) and two olfactory sensitivity tests. For the PIT, twenty hand drawn colour pictures objects were selected from Spanish familiar products. For the odour threshold test (OTT) one plain distilled water flask and six different n-propanol dilutions, using twice-distilled water, were used. The concentrations, in percentage of n-propanol in water, were: 0, .5, 1, 2, 3, 4 and 5 %. The dilutions were placed in bottles with opening diameters of 2.5 cm. The odour identification test (OIT) consisted of identifying the six odours that had been correctly identified by all subjects when considering the total sample of 20 drawings of the PIT. Subjects were blindfolded in order to prevent visual identification of some of the odorants.

Procedure

Odour thresholds (OT) were calculated from the mean of the first stimuli perceived in the ascending staircases or the last stimuli perceived in the two descending staircases of the six stimuli concentrations and the plain water flask. Thresholds were defined as this mean value, considering the four staircases. For the olfactory identification test (OIT) the six stimuli chosen from the PIT results were: lilac bath gel, baby powder, paint, olive oil, chocolate and tomato juice. The subjects had to choose from four pictures, one of which was the correct odour. All of the four pictures were presented to the subjects in the same order and each subject chose the picture that corresponded to the odour. The cognitive impairment tests (MMSE and Eurotest) were answered, counterbalanced by subjects, one previous to the olfactory experiments and the other at the end.

Results and Discussion

Three types of statistical analyses will be used to obtain our results: (1) two backward regression analysis including all quantitative and binary independent variables and considering MMSE and Eurotest scores as dependent variables, respectively. (2) significant linear correlations considering all quantitative and binary variables [age **(A)**, gender **(G)**, years of schooling **(YS)**, number of correct pictures identified **(NCPI)**, olfactory threshold **(OT)**, number of correct odour identifications (**NCOI**), **MMSE** scores and **Eurotest** scores] and (3) Answer tree results considering as dependent variables the MMSE and the Eurotest, respectively, and as independent variables the number of correct olfactory identifications, olfactory thresholds, number of correct pictures identified, age, years of schooling and civil status. The first two analyses are parametric and the third one is the non-parametric one. AnswerTree creates a classification systems displayed in decision trees. The growing method of our trees used CHAID. It uses chi squared statistics to identify optimal splits.
To study the relevant variables that explain the MMSE scores or Mini-mental estimated scores (MMES), a backward standardized regression equation was calculated (using SPSS/PC+, Version 12). The standardized model includes the MMSE scores as dependent variable and all the other variables, except the Eurotest scores, as independent ones. After three iterations three independent variables were removed (OT, G and YS) and the remaining equation was:

$$MMSE` = \beta_1 \, NCOI + \beta_2 \, NCPI + \beta_3 \, A \qquad (1)$$

where: MMSE´ = Estimated Mini-mental score ; NCOI = Number of correct odor identifications ; NCPI = number of correct picture identifications; A= age.

The backward regression method obtains the significant coefficients for each variable. The significant standardized regression equation coefficients for each independent variable are:

$$MMSE` = 0.451 \, NCOI + 0.435 \, NCPI - 0.222 \, A \qquad (R^2 = 0.686) \qquad (2)$$

68% of the variance of Mini-mental scores is explained by NCOI, NCPI and A. The Mini-mental scores can be predicted, mainly from the number of correct olfactory identifications, then by the number of correct picture identifications and, finally by age in a negative sense. This final aspect means that as subjects age, the Mini-mental scores decreases and cognitive impairment increases. The biggest weight in predicting cognitive impairment, using the MMSE, is found in olfactory sensitivity, measured by the number of correct odour identifications.

Considering the next cognitive impairment test, the Eurotest, to study the relevant variables that explain its scores, another backward standardized regression equation was also calculated. Now the standardized model includes the Eurotest scores as dependent variable and all the other variables, except the MMSE scores, as independent ones. After five iterations five independent variables were removed (OT, G, YS, A, NCPI) and the remaining equation was:

$$Eurotest` = \beta_1 \, NCOI \qquad (3)$$

where: Eurotest´ = Estimated Eurotest score ; NCOI = Number of correct odour identifications.

The significant standardized regression equation coefficients for each independent variable are:

$$Eurotest` = 0.445 \, NCOI \qquad (R^2 = 0.198) \qquad (4)$$

19.8 % of the variance of the Eurotest is explained by NCOI. The Eurotest scores can be predicted from the number of correct olfactory identifications. For predicting cognitive impairment, using the Eurotest, olfactory sensitivity, measured by the number of correct odour identifications, is the only relevant variable.

The second statistical analysis is the correlation matrix with only the significant coefficients.

Table 1. Correlation matrix (Significant values for alpha=0'01, alpha=0'05*)**

	Age	Gender	YS	NCPI	Threshold	NCOI	MMSE	Eurotest
Age	-							
Gender	-0.318**	-						
YS	-	-	-					
NCPI	-0.335**	-	0.263*	-				
Threshold	-	-	-	-	-			
NCOI	-0.260*	-	-	0.325**	-0.530**	-		
MMSE	-0.484**	-	0.258*	0.656**	-0.382**	0.650**	-	
Eurotest	-	-	-	-	-0.403*	0.445*	0.627**	-

Studying first the significant correlations associated to our dependent variables, the MMSE and the Eurotest, one finds: (1). Both tests correlate significatively, for alpha=0'01. This could mean that, since both test are supposed to measure cognitive impairment, each one can validate the other (39.3% of the variance of one of them, is explained by the other). (2). The MMSE is influenced by non-olfactory factors as age, years of schooling and the number of correct pictures identified. Older subjects get lower scores could mean that older subjects have more cognitive impairment. Education level modulates or influences the scores obtained, which is a non wanted influence. The number of correct pictures identified is the highest relationship obtained between the variables studied. Nevertheless, both olfactory measures correlate with this test and in the correct direction (the higher test scores significatively relate with low thresholds, and with high scores in olfactory identification). (3). The Eurotest only relates significatively with both odour measures and in the right direction: negatively with threshold and positively with the number of correct odour identifications. Another positive aspect is that the test does not relate significatively with any of the "disturbing" variables such as age, gender and years of schooling. (4). Another relevant correlation is the one that relates both olfactory measures. The subjects responses for those measures go in the adequate direction: the lowest the threshold measure, the highest the number of correct odour identification. (5). The years of schooling is also related with the number of correct pictures identified and (6). The more pictures correctly identified are related to younger people.

Our third data analysis concern a non-parametric technique: AnswerTree. AnswerTree diagrams use your data to obtain rules that will let you classify cases with maximum accuracy. The first tree (Figure 1) considered the Mini-mental scores as the dependent variable and age, gender, civil status, years of schooling, number of pictures correctly identified, olfactory thresholds and number of odours correctly identified are the independent ones. Among all the independent variables considered, only the number of odours correctly identified can predict Mini-mental scores.

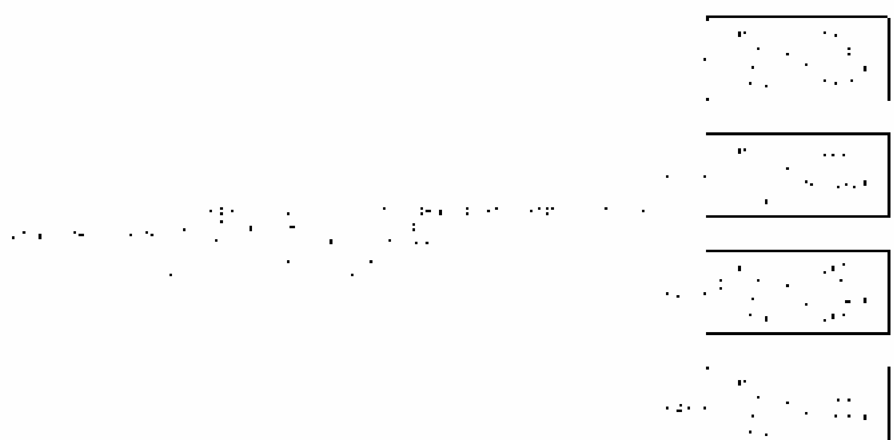


Figure 1. Number of correct odour identifications and Mini-mental scores.

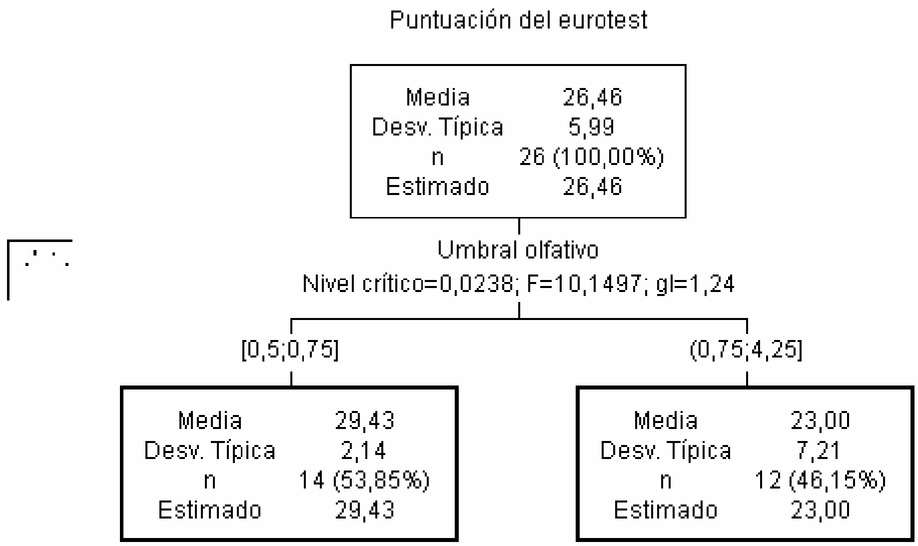


Figure 2. Olfactory thresholds and Eurotest scores.

Zero correct responses can predict the lowest Mini-mental score (Mean value of 19.67), one correct response predicts a mean Mini-mental score of 22.23, two or three correct identification responses predicts a mean Mini-mental score of 24.56 and four or five correct olfactory identifications predicts a mean Mini-mental score of 28.08. As the number of olfactory hits increases cognitive impairment decreases or Mini-mental scores increases. The categories that have been grouped imply that the distribution of Mini-mental scores for two and three correct responses are not different, so they are grouped together. The same thing happens with four and five correct responses.

Considering the Eurotest scores (Figure 2) as the dependent variable and the same independent variables as the ones mentioned above, one can see that the olfactory threshold can predict eurotest results. High thresholds (from a little bigger than 0.75 to 4.25) have a low mean Eurotest score of 23. Low thresholds have a high mean Eurotest score of 29.43. Both groups are evenly split (53.85% vs 46.15%).

Taking into consideration all our results, we can expect, with a high probability, that olfactory data from elderly people that are healthy and non-institutionalized, can predict cognitive impairment either using the Eurotest or the Mini-mental State Examination. Nevertheless, different techniques shed light to different olfactory measures. Using a non-parametric technique (Answer Tree), thresholds can predict cognitive impairment using the Eurotest and the number of correct odour identifications can predict the impairment using the MMSE. Using a parametric technique, backward regression, both the Eurotest and the MMSE are predicted by only one of the olfactory sensitivity measures, the number of correct odour identifications. Both tests are highly positively correlated and, as a consequence, they co-varied in the same direction. Also both olfactory sensitivity measures are significatively related. Therefore olfactory sensitivity measures can predict cognitive impairment and this can lead us to observe that odour impairment can be an early symptom of cognitive impairment.

References

Buck, L. & Axel, R. (1991). Odorant receptors and the organization of the olfactory system. *Cell, 65,* 175-187.

Cain, W.S. (1981). Educating your nose. *Psychology Today, 15,* 7, 48-56.

Carnero Pardo, C. (2005). *El Eurotest: Test Europeo de Detección de Deterioro Cognitivo.* Unpublished doctoral dissertation, Universidad de Granada, Granada.

Carnero Pardo, C. & Grupo Eurotest. (2002). The money test. *Neurology, 58,* 7: A 274.
Carnero, C., , Montoro, T., Barquero, M. S., Frank, A., Mola, S., Gurpegui, M., Manubens, J. M. & Torrecárdenas, H. (on behalf of the TRANS-EUROTEST GROUP) (2005). *Eurotest's diagnostic accuracy in general neurologic clinic.* Poster presented at the annual meeting of the American Neurological Association Meeting, Miami.
Chan, A., Tam, J., Murphy, C., Chiu, H. & Lam, L. (2002). Utility of olfactory identification test for diagnosing Chinese patients with Alzheimer´s disease. *Journal of Clinical and Experimental Neuropsychology, 24,* 2, 251-259.
Engen, T. (1991). *Odor Sensation and Memory*. New York: Praeger.
Folstein, M.F., Folstein, S.E. & McHugh, P.R. (1975). *Mini-Mental State Examination* (Spanish version). Renewed 2005 version in Luttz, Florida: Psychological Assessment Resources.
Garriga-Trillo, A. (1985). *Función Psicofísica y Medida de la Sensibilidad Olfativa*. Unpublished doctoral dissertation, Universidad Autónoma de Madrid, Madrid.
Garriga-Trillo, A. (1987). Olfactory Psychophysics: Sensitivity Measures. In E. E. Roskam & R. Suck (Eds.) *Progress in Mathematical Psychology-1* (pp. 343-349). Amsterdam: North Holland.
Garriga-Trillo, A. (2003). La sensibilidad olfativa puede predecir la enfermedad de Alzheimer. In C. García Gallego, L. Quintanilla Cobián, A. J. Garriga-Trillo, S. Fontes de Gracia, M.C. Pérez-Llantada Rueda & E. Sarriá Sánchez (Eds.), *Cuadernos de Prácticas: Diseños de Investigación en Psicología* (pp.25-32). Madrid: UNED.
Mesholam, R.I., Moberg, P. J., Mahr, R.N. & Doty, R.L. (1998). Olfaction in Neurodegenerative Disease: A Meta-analysis of Olfactory Functioning in Alzheimer´s and Parkinson´s Diseases. *Archives of Neurology, 55,* 1, 84-90.
Moberg, P.J. & Raz, N. (1997). Aging and olfactory recognition memory: Effect of encoding strategies and cognitive abilities. *International Journal of Neuroscience, 90(3-4*), 277-292.
Schiffman, S. S. (1983). Taste and Smell in Disease. *The New England Journal of Medicine, 308,* 21, 1275-1279.
Susuki, Y., Yamamoto, S., Umegaki, H., Onishi, J., Mogi, N., Fujishiro, H. & Iguchi, A. (2004). Smell identification test as an indicator for cognitive impairment in Alzheimer´s disease. *International Journal of Geriatric Psychiatry,14,* 727-733.

* The research reported in this article was funded by Grant BSO2003-00819 (Ministry of Science and Technology, Spain) awarded to A. Garriga-Trillo.

MIXING DURATIONS AND SENSORY MODES IN TEMPORAL MEMORY

Simon Grondin, Pierre-Luc Gamache, Marie-Ève Roussel, Marc Pouliot, Marilyn Plourde
Université Laval, Québec, Canada

ABSTRACT

In this experiment, the hypothesis that memory is a critical source of variance in the temporal processing of brief intervals (base durations from .1 to 1 s) was tested. Participants had to categorize intervals as short or long. The number of base durations and interval types (auditory or visual) randomized within blocks of trials varied from one session to another. The results reveal that mixing three base durations during a session, compared to using only one base duration, decreases the accuracy of time judgments. The results are argued to show that maintaining a representation of time within memory is a major source of variance in the processing of temporal information.

Many time perception researchers assume that there is a central timekeeping system on the basis of which time judgments are made. A common idea is that the so-called *internal clock* is made of a pacemaker-counter device. The pacemaker emits pulses according to some distribution property, and it is the accumulation of these pulses that determines the subjective experience of time: The more pulses accumulated, the longer the perceived duration.

The pacemaker-counter device is at the heart the *Scalar Expectancy Theory* (SET) which gained popularity via its information-processing version (Penney, 2003). This information-processing metaphor of SET is composed of three levels of processing: perceptual (the clock), mnemonic and decisional. In addition to the pacemaker-counter device described above, the clock process is made of a switch component that determines the accumulation of pulses in the counter. The purpose of the experiment reported below is to study the properties of the memory component. Jones and Wearden (2004) reported that using two standards in a the generalization phase of a temporal generalization task resulted in more identification errors. On the other hand, Penney et al. (1998, 2000) reported that when auditory and visual signals are mixed within a session, intervals marked with auditory signals are perceived as being longer than intervals marked by visual signals. This finding is consistent with Grondin and Rousseau (1991). Grondin (2005) recently reported that mixing two base durations within blocks of trials affects duration discrimination, but mixing auditory and visually marked intervals influence neither the discrimination level nor the relative perceived duration.

In the present experiment, the memory load was increased by mixing three base durations; two modalities were also used. Six experimental conditions were obtained by combining these two factors. An experimental session was held for each of the three base durations for a given modality, with each session involving many presentations of each interval in order to favor the development of a stable representation in reference memory and set baseline performance levels. It was only after the three sessions were held that the intervals drawn from the various distributions were randomized within a block of trials.

Method

Participants

Twelve 23- to 30-year-old volunteer students at Université Laval, 5 females and 7 males, participated in this experiment. They were paid CAN$72. ($4. per session) for their participation.

Apparatus and stimuli

The intervals to be discriminated were marked by two 20-ms sensory signals (markers), visual or auditory. The visual markers were produced by a circular, red-light-emitting diode (LED: Radio-Shack #276-088) placed about 1 m in front of the participant, subtending a visual angle of about $.57^0$. The auditory markers were 1-kHz tones with a recorded intensity of about 70 dB SPL. The auditory signals were presented binaurally through headphones (Sony MDR-V600).

Each observer was seated in a chair in a dimly lit room and asked to respond either "short" or "long" by pressing the left or the right button, respectively. Adjacent to each button on the response box was a small light used to provide feedback after each trial. All other aspects of the experiment were controlled by a Zenith microcomputer.

Procedure

The single-stimulus method was employed: Each trial consisted of the presentation of one interval. The participant was asked to judge if the time interval between the two sensory signals belongs to the ‘‘short’’ or to the ‘‘long’’ category of a given distribution of intervals around a mid-point value (base duration). A 1.7-s feedback signal was presented 200 ms after the response, followed by a 1-s inter-trial interval. Feedback indicated whether the presented interval (see the next paragraph) was one of the three short intervals (‘‘short’’ category) or one of the three long intervals (‘‘long’’ category).

The experimental conditions were built around two physical variables: 2 modalities (A=Auditory and V=Visual) which delimited the empty intervals, and three base durations, 100, 550 and 1000 ms. Accordingly, there were six kinds of trials: A100, A550, A1000, V100, V550, V1000. The participants were not presented with the base durations themselves. Each one of the base durations was surrounded by six intervals: 80, 88, 96, 104, 112, and 120 ms (at 100 ms); 440, 484, 528, 572, 616, and 660 ms (at 550 ms); and 800, 880, 960, 1040, 1120, and 1200 ms (at 1000 ms).

There were three loading conditions: 1) one of the six kinds of trials per session, entailing six different experimental conditions (Loading Condition 1); 2) three of the six kinds of trials randomized within one session, either A (Loading Condition 2A) or V (Loading Condition 2V), yielding two experimental conditions, and 3) all six kinds of trials randomized: one experimental condition (Loading Condition 3).

Eighteen sessions lasting about 30 to 35 minutes each were conducted, with 5 blocks of 72 trials per session. Within each block, there were 12 repetitions, in a random order, of each of the six comparison intervals in the Loading Condition 1; four repetitions, in a random order, of each interval of the three base-duration distributions in the Loading Conditions 2A and 2V; and two repetitions, in a random order, for each modality and each interval of the three base-duration distributions, in the Loading Condition 3.

The eighteen sessions were divided in 6 sessions for each of the three Loading conditions, Loading Condition 3 (Sessions 13-18) being the same for every participant. Loading Conditions 1 and 2 were divided into two parts. In Sessions 1-3 (and 7-9), participants were conducted in three of the six kinds of trials of the Loading Condition 1, and in Sessions 4-6 (and 10-12) they were conducted into Loading Condition 2. For six participants, Sessions 1-6 and 7-12 involved auditory and visual markers, respectively; and for the other six participants, it was the reversed order. Each of the six participants were conducted, in Sessions 1-3, in one of six possible orders for the three base duration conditions; and the base duration order in Sessions 7-9, for one given participant, was the same as in Sessions 1-3. At the end, each point on the individual psychometric function in each of the eighteen experimental conditions was based on 60 observations.

Data analysis

For each participant and for each of the 18 experimental conditions, a 6-point psychometric function was traced, plotting the six empty intervals on the *x-axis* and the probability of responding "long" on the y-*axis*.

The *cumulative normal distribution* (CND) was fitted to the resulting curves. Two indices of performance were estimated from each psychometric function, one for sensitivity and one for the perceived duration. As an indicator of temporal sensitivity, estimates of one standard deviation (SD) on the psychometric function were determined. Using one SD (or variance) is a common procedure to express temporal sensitivity.

The other dependent variable was the temporal bisection point (BP). The BP can be defined as the *x* value corresponding to the 0.50 probability of "long" responses on the *y*-axis. The observed shift of the BP for different conditions can be interpreted as an indication of differences in perceived duration. Thus, longer perceived durations are reflected by smaller BP values.

Results

Before the mean of individual results are presented, it should be noted that Figure 1 provides a general picture of the results. In this figure, data from all participants are pooled together. Overall, there was an increase in the probability of responding ''long'' as the value of the intervals on the *x* axis rises. This indicates that, in general, participants were able to do the task correctly in spite of the difficulty inherent in the Loading Conditions 2 and 3.

Essentially, four main features can be extracted from Figure 1. Firstly, at 100 ms, the dispersion of data points in all three loading conditions is much narrower with visual signals than with auditory ones. Secondly, for the other two duration ranges, the dispersion of data points is much narrower in both auditory and visual conditions in the Loading Conditions 2 and 3 than in the Loading Condition 1. Thirdly, in the Loading Conditions 2 and 3, there was a clear overall tendency to produce more ''long'' responses for intervals drawn from the 1000-ms base duration than for intervals from the 550-ms base duration, for which much more ''short'' responses were produced. Fourthly, in the Loading Conditions 2 and 3, there were more ''long'' responses for the shortest intervals (800 ms) of the 1000-ms base duration than for the second shortest interval (880 ms).

For the purpose of individual analyses, a psychometric function was traced, as noted earlier, for each participant and for each of the 18 experimental conditions, and the CND was fitted to the resulting curves. For some participants, 5 points instead of 6 were used as there was clear evidence of confusion in assigning a given interval to its proper distribution in the multiple-base duration conditions. As one would deduct from the fourth point raised in the preceding paragraph, the shortest interval of the 1000-ms distribution (in the Loading Conditions 2 and 3) was the most frequently rejected data point.

In the following series of comparisons, a one-way ANOVA with repeated measures was used, except when indicated, to test whether the mean differences between the three Loading conditions were significant. It was decided that the data from the three base durations would be analyzed separately because of the confusion, in the Loading Conditions 2 and 3, between some data points of the 550- and 1000-ms base durations; there was not much confusion, however, in the 100-ms base duration. Also, since it was preferable not to include the data from one of the participants in one condition (for SD at 550 ms in the visual condition), the modalities were analyzed separately.

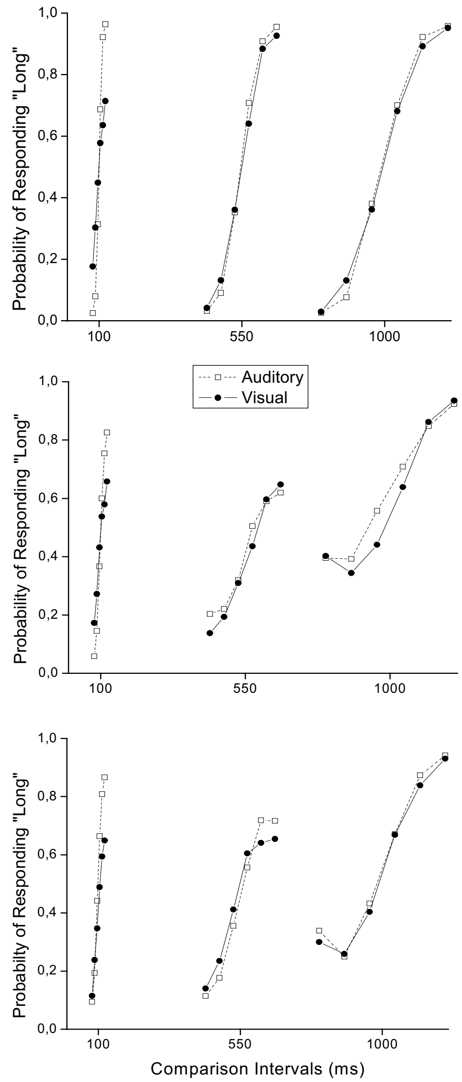


Figure 1. Total probability of responding ‘‘long’’ (grouped data), for auditory and visual signals, as a function of comparison intervals for each base duration condition (on each panel) and for the Loading Conditions 1 (Upper panel), 2 (Middle panel) and 3 (lower panel).

Figure 2 (left panel) shows all the data for SD in the auditory condition. At each base duration, the lowest SD (highest sensitivity) was obtained when only one base duration and one marker type were used during a session. Interestingly, using intervals from three base durations caused at least as much damage in performance as mixing auditory and visual intervals in addition to using intervals from three base durations.

Each ANOVA on SD, at 100 [$F(2,22) = 18.65$, $p < .01$], 550 [$F(2,22) = 9.36$, $p < .01$] and 1000 ms [$F(2,22) = 5.48$, $p < .05$], revealed significant differences. Post-hoc Least Significant Difference (LSD) tests confirmed that SD was significantly higher with the Loading Conditions 2 and 3 than with the Loading Condition 1 (each $p < .01$) at 100 and 550 ms. At 550 ms, the difference between the Loading Conditions 2 and 3 was not significant (p=.072). At 1000 ms, the difference between the Loading Conditions 1 and 2 was significant ($p < .01$), but not the one between 1 and 3 (p=.089).

Figure 2 (right panel) shows the SD for each condition with the visual markers. In the 550- and 1000-base duration conditions, the results resemble those in the auditory condition, with the best performance obtained in the single base duration condition. It should be noted that, at 550 ms, the data of one participant was not used for the analysis because, in the Loading Condition 2V, the SD was 682 ms (R^2 values at .30), which is almost 500 ms more than the next highest SD in this condition. The SD for this participant was 57.7 ms and 181.5 ms in the loading conditions 1 and 3, respectively.

The ANOVA on SD at 550 [$F(1.23, 12.29) = 6.19$, $p < .05$] and 1000 ms [$F(2,22) = 10.61$, $p < .01$] revealed significant differences. Post-hoc LSD tests confirmed that SD was significantly higher in the Loading Conditions 2 ($p<.01$) and 3 ($p<.05$) than in the Loading Condition 1. However, the differences at 100 ms were not significant [$F(2,22)= .60$, $p=.56$].

BP was also calculated in each experimental condition. The question addressed here is whether mixing visual and auditory signals influenced the perceived duration. However, as revealed by Figure 1, in the 550- and 1000-ms conditions, it was not mainly the modality that influenced the probability of responding ''short'' or ''long'', but rather the contrast between these two duration ranges. Therefore, to test the potential modality effect on perceived duration, our analysis focused on the 100-ms data. Figure 3 illustrates the difference between auditory and visual intervals for BP. In the Loading Conditions 1 and 2, the difference between modalities was not significant, ($t(11)=1.35$, $p=.20$) and ($t(11)=.84$, $p=.42$); it was significant, however, in the Loading Condition 3 ($t(11)=2.66$, $p<.05$).

Finally, in order to test the stability of the variability over time, a coefficient of variation (CV) was calculated: SD/BP. A 2 x 3 ANOVA with repeated measures revealed thatthere was a significant modality effect [$F(1,11) = 28.27$, $p < .01$] and a significant base duration effect [$F(1.05,11.56) = 13.08$, $p < .01$]. The significant interaction effect [$F(1.03,11.29) = 14.67$, $p < .01$] indicates that there were no significant differences between base duration conditions in the auditory condition. In the visual condition, however, the CV was significantly higher at 100 ms than in the other two conditions.

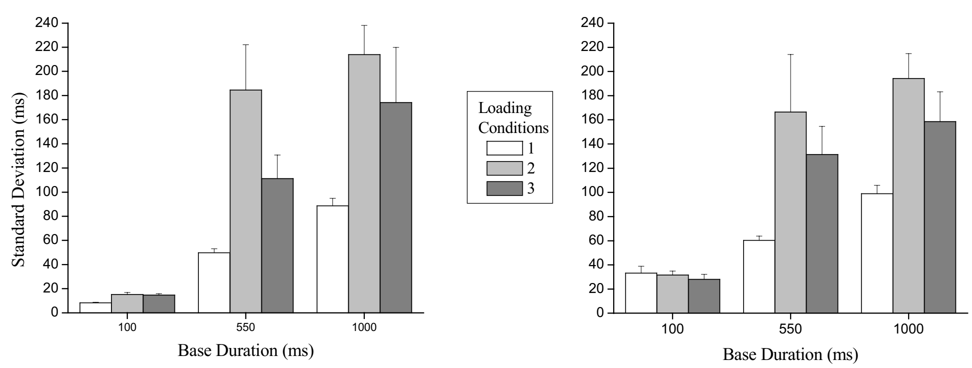


Figure 2. Mean standard deviation, for the auditory-marker condition (left panel) and the visual-marker condition (right panel), in the Loading Conditions 1, 2 and 3 for each base duration (Bars are standard errors)

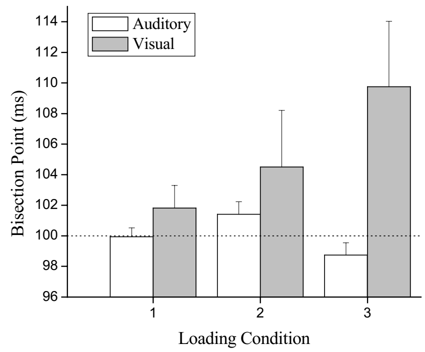


Figure 3. Mean bisection point for 100-ms intervals, in auditory and visual conditions, in the Loading Conditions 1, 2, and 3. (Bars are standard errors)

Conclusion

The main finding of this experiment is that participants were capable of simultaneous timekeeping activities, as can be seen in Figure 1. However, this multiple-timing performance had a cost, as revealed by the increased SD in the Loading Conditions 2 and 3 when compared with the basic level (Loading Condition 1). This finding applies to each base duration (but not in the visual condition at 100 ms). The Loading Conditions 2 and 3, on the other hand, did not differ from one another. Indeed, SD consistently tended to be lower in the Loading Condition 3 than in the Loading Condition 2. This is somewhat surprising given that the task may have been more demanding in the Loading Condition 3, but it is consistent with Grondin (2005).

In regard to the bisection point, the analysis focused on the 100-ms base duration. The results revealed no difference between the auditory and visual conditions in the Loading Conditions 1 and 2, i.e., when there was no direct comparison of visual and auditory intervals within a session. However, the bisection point was significantly lower in the auditory condition than in the visual one in the Loading Condition 3 where base durations were mixed and auditory and visually marked intervals were randomized. In other words, the visual intervals were perceived as being shorter than the auditory ones, a result that is consistent with other recent findings on this issue (Penney & al., 2000), but not with Grondin (2005) where the lowest base duration was 250 ms.

References

Grondin, S. (2005). Overloading temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31* (in press).

Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception and Psychophysics*, *49*, 245-256.

Jones, L. A., & Wearden, J. H. (2004). Double standards: Memory loading in temporal memory. *Quarterly Journal of Experimental Psychology, 57B*, 55-77.

Penney, T. B., Allan, L. G., Meck, W. H., & Gibbon, J. (1998). Memory mixing in duration bisection. In D. A. Rosenbaum & C. E. Collyer (Eds.). *Timing of behavior: Neural, computational, and psychological perspectives* (pp. 165-193). Cambridge, MA: MIT Press.

Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1770-1787.

Penney, T. P. (2003). Modality differences in interval timing: Attention, clock speed, and memory. In W. H. Meck (Ed.). *Functional and Neural Mechanisms of Interval Timing* (pp. 209-234). Boca Raton, FL: CRC.

Acknowledgment. This research was made possible by a grant from the Natural Science and Engineering Research Council of Canada to SG, and Summer Scholarships to PLG and MER from NSERC. We wish to thank Hannes Eisler and Françoise Macar for their comments on this text. (simon.grondin@psy.ulaval.ca)

ANATOMY OF STIMULUS COMPARISON

Åke Hellström
Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
Email: hellst@psychology.su.se

Abstract

Stimulus comparison by human observers is a more complex process than the subtrac-tion carried out by a simple *comparator*. This is shown by the patterns of time- and space-order "errors." Specifically, experiments show that the two compared stimuli exert differential influence on the comparative response. This invalidates models based on additive bias or semantic congruity. The sensation-weighting (SW) model can account for many of the results. One important consequence is that the discriminability of two stimuli depends on which of them is varied.

Life is filled with choices. Choices are based on comparisons between alternatives. One metaphor is the balance, the symbol of justice: Adding to A or taking away the same amount from B yield the same correct answer to "Which is heavier, A or B?". Many psychophysical models contain a *comparator*, which is assumed to work this way, like an electronic comparator device. It may seem self-evident that stimulus comparison by human observers is basically comparison in this sense – to form the difference and check its sign.

The case below illustrates the basic meaning of 'compare' – bring things together and examine their relations. The difference A-B can be observed more or less directly:

A ______________
B _______________

However, the comparison situation below may be more typical:

______________ (TIME- OR SPACE-INTERVAL) _______________

A B

A and B are not literally brought together, and which of them is longer must be inferred by a more complicated mental operation than direct sensing of the difference. Is this operation equivalent to subtraction? If not, how can it be described? This aspect of stimulus comparison, which present-day psychophysics usually ignores, will be the main theme of this paper.

Comparison and discrimination

Stimulus comparison has usually been studied in the context of discrimination. In the *Method of Constant Stimuli,* one stimulus, usually the first, is kept constant (standard, St), and the other (comparison stimulus, Co) is varied in order to determine the *just noticeable difference* (JND) and the *point of subjective equality* (PSE). PSE usually differs from St, and this difference is called the Constant Error (CE). With a temporal interstimulus interval (ISI), CE is called *time-order error* (TOE), with a space-interval, *space-order error* (SOE) (Fechner, 1860).

Rather than measuring jnd's, researchers today tend to use the two-alternative forced choice (2AFC) paradigm, and interpret the results in terms of signal-detection theory, which only acknowledges sensitivity (d') and bias. Order effects are treated as biases and "eliminated" by averaging the results from the two orders of presentation. Thus TOEs are defined away and never observed. Still, they are sometimes rediscovered: Klein (2001) noted an "interval bias" (favoring the second interval, i.e., a negative TOE) in 2AFC contrast discrimination data, and showed that this bias affects the value of d' if it is not taken into account in the computations.

Modeling order effects

Koehler & Ridpath (1982) accounted for ordinal comparisons in terms of an *extended Bradley-Terry-Luce model*, $p(\text{A>B}|\text{A,B}) = \psi(\text{A})/[\psi(\text{A}) + w\ \psi(\text{B})]$, where $\psi(\text{A})$ and $\psi(\text{B})$ are are stimulus magnitude parameters and w is an order effect parameter. The order effect is additive in ln ψ: $\text{logit}[p(\text{A>B})] = \ln \psi(\text{A}) - \ln \psi(\text{B}) - \ln w$. Additive effect models are often associated with the notion that the TOE is a response bias and has little or nothing to do with perception. Such explanations were made problematic by results such as those of Woodrow (1933) and Needham (1935), who found that with a roving standard, the size and direction of the TOE, and its dependence on the length of the ISI, changed with the magnitude level of the standard. John (1975) suggested that such effects are due to an implicit verbal response, making the observer more prone to respond, for instance, "Co is louder than St" for loud than for soft sounds. Although such *semantic congruity effects* do affect response times as well as accuracy (Petrusic & Baranski, 1992), this and other response-related factors seem to be ruled out as explanations for the TOE by the results of Jamieson and Petrusic (1975) and Hellström (1977), who showed that varying the response mode does not substantially affect the TOE or its dependence on the stimulus level.

Describing the TOE as an additive effect was invalidated by the important study of Michels and Helson (1954), using lifted weights and a rating scale for the comparative judgment (J) to measure the subjective impression of the second stimulus in relation to the first. They presented the two stimuli in both time-orders, and found that the effect on J from varying Co was much greater when Co was last in the pair than when it was first. Thus the two stimuli had differential effects on J, which means that the comparison could not be based on simple subtraction of the sensation magni-tudes. This also meant that the TOE was usually different in the two presentation orders, so that it could not be handled by an additive term.

In Michels and Helson's model, a comparative judgment (J) of the second of two successive stimuli in relation to the first is the difference between the magnitudes of the second stimulus and that of its *comparative adaptation level* (CAL), which is based, as every AL, on the pooling of past stimulus magnitudes. Specifically, CAL is formed by weighting together the magnitudes of the first stimulus and of the series AL (SAL) in the proportions s and $1\text{-}s$. Simplifying from the original account, with the presentation order St-Co, $J = \text{Co} - s\ \text{St} - (1\text{-}s)\ \text{SAL}$; with the order Co-St, $J = \text{St} - s\ \text{Co} - (1\text{-}s)\ \text{SAL}$. In the order Co-St, the net effect on J from a change in Co is thus lower, by the factor s, than in the order St-Co. This explanation is equivalent to assimilation of the first stimulus towards the general background, although applied not only to standard stimuli on different levels (Needham, 1935; Woodrow, 1933), but to stimuli within a single series.

Michels and Helson's view was based on the common notion that the comparison of two successive stimuli occurs between the directly perceived second stimulus and the recalled image of the first stimulus. However, this seems to be a simplified description of what happens (cf. Fernberger, 1919). For one thing, the responding is often not completed while the second stimulus is still present. Thus, each stimulus may interact in some way with the contextual stimuli.

Fechner (1860, vol. 2, p. 141) also pointed out that the two stimuli may interact: "In fact, an influ-ence of the temporal position could very well be due to the fact that the second stimulus impinges on a sense organ that is already changed by the first stimulus, insofar as on one hand a certain pro-longed effect of every stimulation occurs, on the other hand a blunting due to every stimulation; these are effects that work in opposite directions, and their conflicts and respective dominance in various circumstances could explain the Proteus-like variation of the time-order error across condi-tions, which I have noticed in my weight and touch experiments." [Translated by Å. Hellström and H. Eisler]

Modeling developed

With s<1, as is implicit in thinking based on pooling or assimilation, Michels and Helson's model can only account for cases where the effect of the first stimulus on the judgment is less than that of the second, and the effect of the second stimulus varies little between conditions.

These limitations should cause few problems in the case of lifted weights (Hellström, 2000). However, Hellström (1979), in an experiment with tone loudness with widely varying temporal conditions, for brief stimuli and ISIs, used an "equal" category and a Thurstonian scaling method based on response proportions. With short stimuli and ISIs he found a much higher effect on the scaled difference from the first than the second stimulus. Hellström found that the differential stimu-lus effect varied greatly betwen conditions, occurred over small as well as large stimulus ranges, and could account for the effect of the stimulus level on the TOE. It was also clear, considering the en-tire set of results, that centering or adaptation was involved. To account for these results, Hellström designed the *sensation-weighting* (SW) *model*, which generalizes Michels and Helson’s model:

$$d = k\,\{[s_1\,\psi_1 + (1 - s_1)\,\psi_{r1}] - [s_2\,\psi_2 + (1 - s_2)\,\psi_{r2}]\}, \qquad (1a)$$

where d is the scaled subjective difference, k a scale constant, ψ_1 and ψ_2 the sensation magnitudes of the stimuli, s_1 and s_2 weighting coefficients, and ψ_{r1} and ψ_{r2} the subjective magnitudes of the reference levels (ReLs). There are no restrictions on the s values. ReL generalizes Helson's AL. The two ReLs may differ (this made possible the estimation of "absolute" weights [s values] in Hellström [1979]). Each stimulus magnitude is weighted together with its ReL, and the weighted compounds are subtracted from each other. A bias term may be added if needed.

TOE as explained by SW model

TOEs, describable by the SW model, occur for physical magnitudes (Hellström, 2003a) as well as for aesthetic preferences (Hellström, 2001; Koh, 1967). TOEs are often negative (first stimulus underestimated relative to second), more so the longer the ISI. Practice effects may reverse this relation. Hellström (1979, 1985, 2000) ascribed these results to a combination of (a) $s_1 < s_2$, (b) ReLs lower than the average stimulus, and (c) changes in ReLs with practice.

In Hellström's studies (e,g., 1979; 2003a), the interindividual variation in the TOE was much larger than the variation in the weights. This suggests that the TOE is not the primary phenomenon, but instead a by-product of the comparison process. From the SW model (setting $\psi_{r1} = \psi_{r2} = \psi_r$) it follows that for a pair of physically equal stimuli with magnitude ψ, TOE = $k\,(s_2\text{-}s_1)\,(\psi\text{-}\psi_r)$. Thus the TOE is proportional to the product of the weight difference and the ReL. Simultaneous variation in these factors may cause the "Proteus-like variation" of the TOE.

Weighting and discriminability

It is obvious that if the TOE is not taken into account, the measured jnd will be biased, and will vary with the stimulus level as the TOE does. This will also affect d' in 2AFC (Klein, 2001). However, the differential weighting has a still more important consequence: To make a discriminability mea-sure (e.g., jnd, or the stimulus change that yields $d' = 1$) meaningful, it must be specified which stimulus is changed, the first or the second. Hellström and Rammsayer (2004) found, using an adap-tive discrimination method for noise bursts in the 50- and 1000-ms range, different jnd's with the orders St-Co and Co-St. Rammsayer and Wittkowski (1990) obtained, for comparisons of successive time-intervals, a *constant position error* – a higher proportion of correct responses with St-Co than with Co-St. Hellström and Rammsayer showed this to be an effect of the weight relation $s_1 < s_2$.

What is the process?

There are at least two, mathematically equivalent, process interpretations of the SW model: (a) Two weighted compounds are formed and subtracted from each other, as in Equation 1a. (This idea builds on Helson's pooling concept.) (b) The weighted difference between each stimulus and its ReL is formed, then the simple difference between these differences is formed and the difference between the ReLs, if any, is added, as in Equation 1b:

$$d = k\ \{[s_1(\psi_1 - \psi_{r1})] - [s_2\ (\psi_2 - \psi_{r2}) + (\psi_{r1} - \psi_{r2})]\}, \qquad (1b)$$

Why differential weighting?

Perception specializes in sensing changes in stimuli and stimulus relations. This must be because these changes carry the most information about our environment. Egon Brunswik, in his ecological psychology, emphasized that the goal of perception is to maximize the correlation between the distal stimuli in our ecology and our perception of them (i.e., the perceived change resulting from a stimulus change). This is done by weighting each of the cues in proportion to its ecological validity.

Hake et al. (1966), formulated Brunswik's basic idea as "noise reduction in perception." In this vein, we may think of the weighting in stimulus comparison as optimizing discrimination in the sense of maximizing the signal-to-noise ratio (SNR) of a sudden change in the relation between the stimuli. A test variable for changes in this relation is set up and monitored. But instead of using ψ_1-ψ_2 from a simple "comparator," the observer uses a different test variable, as defined by Equations 1a and 1b. It is biased, leading to systematic errors such as TOEs, but the bias is negligible near the AL, which by definition should be close to the current level of stimulation. When s_1 and s_2 are chosen optimally, this test variable yields a higher SNR than simple subtraction. Simulations (Hellström, 1989) suggest that this gain can be considerable.

The optimal weights for the stimuli are inversely related to the uncertainties of stimulus informa-tion. The resulting sharpening of discrimination must indeed have survival value, which suggests that the weights mirror the stimulus comparison process and also partially compensates for loss of stimulus information from, for instance, cognitive deterioration. They may also mirror different comparison strategies. The ReLs may be used in an assimilative (s<1) as well as a contrastive (s>1) manner, and the optimal s values depend on standard deviations and correlations. The general idea is that specific stimulus information is supplemented with other, usually more generic, information, which is used as a "crutch."

This optimization idea is similar to that of Huttenlocher et al. (2000), who described judgments of past single stimuli using a model with a weighted compound of the magnitude of the stimulus and that of a prototype. They worked out equations for optimal weights, in terms of minimizing the error magnitudes, as a function of the uncertainty of the memorized stimulus. When this uncertainty increases, the optimal weight for the stimulus goes down and that of the prototype goes up.

Evidence for discrinination optimization

The evidence for discrinination optimization in stimulus comparison is rather indirect so far, but it includes the following observations:

(i) Particularly for brief stimuli (e.g., Hellström, 1979, 2003), a striking transition from $s_1 > s_2$ for short ISIs to $s_1 < s_2$ for long ISIs is observed. It is particularly s_2 that changes, so that the average weight magnitude (and thus stimulus discriminability) increases with longer ISIs. This pattern suggests that the weights reflect the shift from stimulus interference to memory loss.

(ii) For patients with solvent-related brain dysfunction (Hellström, Åslund, & Almkvist, 1985) the weight for the first of two compared successive line-lengths decreased for longer ISIs, in accordance with the impaired visual memory found with neuropsychological tests.

(iii) With lateralized presentation of simultaneous visual stimuli (lines), there is a higher weight for the left stimulus (Hellström, 2003; cf. Masin & Agostini, 1991). This seems to be another manifestation of the more efficient processing of visual information in the right hemisphere.

(iv) In Hellström (2004), participants judged which one of two successive musical triads, varying in their middle note, sounded more in the direction of major. Stimulus weights as well as majorness scale values for the triads were fitted for each participant. In a still unpublished analysis, a principal-component analysis of the raw weight differences B_1-B_2 for the four ISIs yielded two large compo-nents, which were varimax rotated. The scores for the first component correlated significantly with the scaling ability (slope of majorness value vs. Hz value of middle note). For good and average scalers, the weights for the two triads were nearly equal, but B_1 went down for the longest ISI, 3200 ms, suggesting some memory loss. For weak scalers B_2 greatly exceeded B_1, suggesting that they relied mainly on the second triad and had difficulty remembering the first one. Thus, the weight patterns mirrored the individual comparison strategies, which in turn depended on the capacity to attend to and remember the relevant features of the triads.

Anatomy of stimulus comparison

Given its simplicity, the SW model does a good job of accounting for data (Hellström, 2003) and seems to reveal important features part of the anatomy of comparison. However, it is descriptive and says nothing about the process that leads up to the comparative judgment. *Random-walk diffusion models* (Link, 1992; Usher & McClelland, 2001) address this process, incorporating response times in the domain of investigation, by assuming that the judgment occurs when enough evidence has accumulated that favors one alternative over the other. Existing models assume that the decision-making mechanism uses a simple comparator which just performs subtraction. Changing this assumption to allow for differential weighting seems like a promising possibility.

Fernberger (1919) investigated the stimulus comparison process for various kinds of stimuli using a phenomenological method. This method did not seem to reveal much about the underlying mechanism, at least not in terms of hints on weighting effects. However, one interesting result was that even when comparing sound or light intensities, the observers frequently reported kinesthetic sensations. Fernberger remarked, "It is a curious fact that all forms of sensation should be transposed into kinaesthetic terms, – that kinaesthesis should become the "common denominator" of all the other modalities of sensation" (p. 150). This suggests that one strategy may be to translate the sensa-tion magnitudes into another, more easily used modality. It also suggests that stimulus comparison is a general cognitive process (Hellström, 1985) and not specific to the particular sense modalities.

Conclusion

We are still far off a complete account of what happens when two stimuli are compared by a human observer. However, looking for the relevant phenomena should help advance our knowledge. As should be clear from the above, a theory that ignores the differential effects of the stimuli on the comparative judgment will never come near the truth, regardless of its level of sophistication.

References

Fechner, G. T. (1860). *Elemente der Psychophysik* [Elements of psychophysics]. Leipzig, Germany: Breitkopf & Härtel.

Fernberger, S. W. (1919). An introspective analysis of the process of comparing. *Psychological Monographs, 26* (117), 1-161.

Hake, H. W., Rodwan, A., & Weintraub, D. (1966). Noise reduction in perception. In K. R. Hammond (Ed.), *The psychology of Egon Brunswik*. New York: Holt.

Hellström, Å. (1977). Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research, 39*, 345-388.

Hellström, Å. (1979). Time errors and differential sensation weighting. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 460-477.

Hellström, Å. (1985). The time-order error and its relatives: mirrors of cognitive processes in comparing. *Psychological Bulletin, 97*, 35-61.

Hellström, Å. (1989). What happens when we compare two stimuli? In G. Ljunggren & S. Dornic (Eds.), *Psychophysics in action*. Berlin: Springer.

Hellström, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 6-17.

Hellström, Å. (2001). Time-order effects for aesthetic preference. In G. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Fechner Day 2001* (pp. 421-426). Lengerich: Pabst.

Hellström, Å. (2003). Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics, 65, 1161-1177.*

Hellström, Å. (2004). Majorness comparisons of triads. In A. M. Oliveira, M. P. Texeira, G.F. Borges, & M. J. Ferro (Eds.), *Fechner Day 2004. Coimbra, Portugal: ISP* (pp. 389-394).

Hellström, Å., Åslund, U., & Almkvist, O. (1985). Comparing computer-presented successive stimuli: Indications of cognitive malfunctioning in solvent-exposed workers. In *Neurobehavioral Methods in Occupational and Environmental Health*. Extended abstracts from the second internatio-nal symposium, Copenhagen, 6-9 August 1985 (Environmental Health Series, Document 3; pp. 115-119). Copenhagen: WHO.

Hellström, Å., & Rammsayer, T. H. (2004). Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psycho-logica, 116,* 1-20.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*, 220-241.

Jamieson, D. G., & Petrusic, W. M. (1975). Presentation order effects in duration discrimination. *Perception & Psychophysics, 17*, 192-207.

John, I. D. (1975). A common mechanism mediating the time-order error and the cross-over effect in comparative judgments of loudness. *Australian Journal of Psychology, 27*, 51-60.

Klein, S. A. (2001). Measuring, estimating, and understanding the psychometric function: A commentary. *Perception & Psychophysics, 63*, 1421-1455.

Koehler, K. J., & Ridpath, H. (1982). An application of a biased version of the Bradley-Terry-Luce model to professional basketball results. *Journal of Mathematical Psychology, 25*, 187-205.

Koh, S. D. (1967). Time-error in comparisons of preferences for musical excerpts. *American Journal of Psychology, 80*, 171-185.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hove and London: Erlbaum.

Masin, S. C., & Agostin , A. (1991). Attentional scanning and space errors. *Perception & Psychophysics, 50*, 285-289.

Michels, W. C., & Helson, H. (1954). A quantitative theory of time-order effects. *American Journal of Psychology, 67*, 327-334.

Needham, J. G. (1935). The effect of the time interval upon the time error at different intensive levels. *Journal of Experimental Psychology, 18*, 539-543.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 962-986.

Rammsayer, T. H., & Wittkowski, K. M. (1990). [Time-order error and position effect of the standard stimulus in the discrimination of short durations]. *Archiv für Psychologie, 142*, 81-89.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, compe-ting accumulator model. *Psychological Review, 108*, 550-592.

Woodrow, H. (1933). Weight-discrimination with a varying standard. *American Journal of Psychology, 45*, 391-416.

A STUDY ON THE IMPACT OF INSTRUCTION ON THE RESPONSE STRATEGIES EVOKED IN PARTICIPANTS

Joeri Hofmans, Walentina Cools, Pieter Verbeke, Nico Verresen, & Peter Theuns

Vrije Universiteit Brussel, Faculty of Psychology and Education
Research Methods and Psychometrics
joeri.hofmans@vub.ac.be

Abstract

In earlier research on the influence of the orientation of the verbal qualifiers on an 'agreement' scale, Hofmans, Baekelandt, Cools and Theuns (2004) found a peculiar pattern of response behaviors among their participants. The authors found that most (72%) participants' subjective null-point coincides with the 'fully disagree' label, whilst others place their subjective null-point on the middle category, 'neutral' (28%). Analyzing the data for these two groups separately appeared useful. The participants whose subjective null-point coincides with 'fully disagree', are less unanimous about the intensity of the labels of a decreasing scale (e.g. fully agree – rather agree – neutral – rather disagree – fully disagree) than they are about the labels of an incremental scale (e.g. fully disagree – rather disagree – neutral – rather agree – fully agree). No such difference was found in the middle null-point group. Here, it makes no difference whether the labels are placed on an incremental or a decreasing scale.

This paper discusses three experiments based on the cross modality matching paradigm (Stevens, 1966). The first experiment was originally designed to produce scale values for different verbal qualifiers on an evaluation-scale. The second experiment is a sidetrack of the experiment by Hofmans et al. (2004) and investigates how instructions impact upon the response strategies found in their research. The third experiment is set up specifically to investigate the two discerned response strategies in depth. Here the first concern is whether the two response strategies will be reproduced? The second research question is whether similar two response strategies occur when different response dimensions are used. As Hofmans et al. (2004) found these strategies for an agreement scale, it could be interesting to explore other commonly used itemized rating scales. Probably some dimensions are more prone to this phenomenon than other, moreover, participants may not use a single kind of response strategy for all dimensions.

Experiment I

In questionnaires, rating scales constitute the most common response modality (Poulton, 1989). Many different formats of rating scales exist, but probably the most frequently used format are itemized rating scales where each response option has its specific verbal qualifier (Breakwell et al., 2000). Unfortunately, rating scales are prone to several biases (Poulton, 1989; Cools, 2004). Therefore the prime goal of this research was to construct an evaluation-scale that would be convenient and yet meet major psychometrical criteria. A first concern in fulfilling these psychometrical criteria was to find out whether some common agreement on the quantification of the verbal qualifiers exists. If such agreement concerning the perceived meaning (and thus the relative quantitative value) of the qualifiers exists between participants, then we also want to establish a scale value for each verbal qualifier.

Method

Participants were 28 males and 121 females between 17 and 33 years old (average age = 20,7). None of the them had experience with experiments on psychophysical scaling. The experiment was run in a laboratory setting on a personal computer with a 17 inch monitor and a resolution of 1024*768 pixels.

For validation of the responses, and to give the respondents an opportunity to practice, their first task consisted of a calibration run in which they had to make 10 numerical estimations of line lengths and draw 10 lines to represent numbers (see also: Lodge, 1981). Next, the participants were presented with 22 verbal qualifiers from the 'evaluation' dimension (e.g. 'bad', 'good', 'average',…), and they had to give numerical estimates of the perceived magnitude of these verbal qualifiers (modulus = 'so – so', value = 40). Finally, the participants were asked to draw line lengths proportional to the perceived magnitude of these verbal qualifiers (modulus = 'so – so', value = 60 pixels). Throughout the whole experiment (except for the calibration run) this standard stimulus or modulus was used and it was visible on the screen all the time.

Results and discussion

Although this research was not set up with the purpose to elicit the two response strategies, they nevertheless appeared in our data. For reasons of clarity, only 5 of the 22 verbal qualifiers are reported here, although a similar strategy is found for the other qualifiers too. There were 44 (29,5%) participants who associated their 'subjective null-point' with the midmost category (NPM), whilst 105 (70,5%) participants considered the label 'extremely bad' as their 'subjective null-point' (NPL).

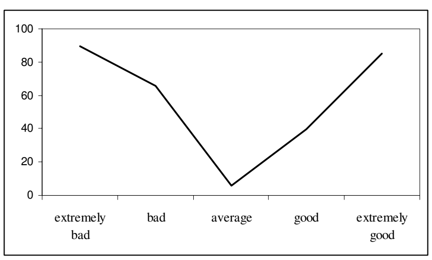


Figure 1: average ratings for respondents following an NPM strategy

Figure 2: average ratings for respondents following an NPL strategy

A possible explanation for this phenomenon can possibly be found in the distinction between one-dimensional and bidimensional scales. People who follow an NPM strategy may perceive the evaluation scale to be bidimensional. So, for these participants 'extremely bad' is the counterpart of 'extremely good'. Other respondents, who follow the NPL strategy, may consider 'extremely bad' to signify 'absence of goodness', and therefore perceive the scale to be one-dimensional.

If this argumentation holds, then the dimensionality of the scale is at least partly 'in the eye of the beholder' and not just some objective property of the scale. Since it was our intention to derive scale values for each verbal qualifier, the existence of these two response strategies complicates our task. A possible solution to this problem would be to use a different algorithm, depending on the response strategy used, when deriving the scale values. The algorithm would be such that, people with an NPM strategy would contribute to the scale values as if they responded following the NPL strategy. However, at least two problems are related with such approach. First of all, such post-hoc method does not assure that the (NPM-) respondent would have produced the same scale values if he or she were asked to respond using an NPL strategy. Secondly, as shown by Hofmans et al. (2004), possibly valuable information can be lost when recalculating the participants' judgements. A different approach would be to manipulate the response strategy through specific instructions. However, before one can start instructing the participants to respond according to some certain strategy, the cause of the difference in strategies must be well known. The fact that the two response strategies are retrieved in two independent studies confirms the supposition of their validity.

Experiment II

If the 2 different response strategies emerge due to a different perception of the dimensionality of the scale, then the use of a standard cross-modality experiment may be problematic since the respondents are instructed to think in ratios, thereby producing a ratio scale. This however implies that only positive numbers can be used, suggesting an NPL strategy or suggesting that the scale to be judged is one-dimensional. The goal of this experiment was to see whether the participants shift strategies when also allowed to make judgements with negative numbers.

Method

This experiment used the same participants as Hofmans et al. (2004), namely 5 men and 31 women with an average age of 23,5. None of these participants had experience with psychophysical scaling experiments. The experiment was run in a laboratory setting on a personal computer with a 17 inch monitor and a resolution of 1024*768 pixels.

After a calibration phase similar to the one described in experiment 1, the participants were asked to assess the intensity of 5 verbal qualifiers of an agreement-scale. As in experiment 1, the respondents assessed the verbal qualifiers using the original numerical estimation (modulus = agree, value = 40), thus using only positive numbers, and line length production (modulus = agree, value = line of 202 pixels). In addition, we added another condition where we allowed the participants to use negative numbers when using numerical estimation.

Results and discussion

Seven of the 37 participants made use of the possibility to answer with negative numbers. 6 of the 7 participants shifted from a NPL strategy to a strategy using negative numbers (see figure 3). Of the 8 participants who followed the NPM strategy, one used negative numbers.

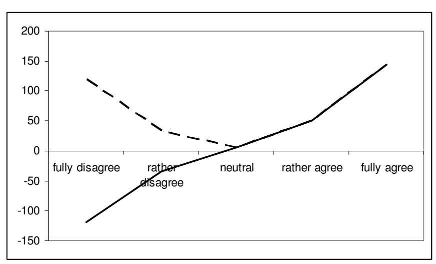


Figure 3: average ratings for respondents who changed their strategies
Note: the dashed line represents the absolute values of the average ratings.

How can one clarify this shift from an NPL to a strategy similar to the NPM strategy? A possible explanation lies in the fact that the instructions in a standard cross-modality experiment can indeed be limiting for some participants. Since they are instructed to think in ratios when evaluating the stimuli, they are unintentionally pushed into a NPL strategy. When we cancel this restriction, for example by allowing negative numbers, some people change strategy. It has to be mentioned that this is only one possible explanation and that we do not introduce it as something impeccable.

Experiment III

As already mentioned, this research was designed to replicate the response strategies (NPL and NPM) and to estimate their magnitude. Also the impact of the manipulation of the instructions was investigated in this study.

Method

The experimental group consists of 24 participants, 17 females and 7 males between 21 and 67 years old (average age = 33). None of the participants had experience with psychophysical scaling experiments.

The experiment is a paper & pencil test developed in MS PowerPoint and printed out on 101 cards in A6 format. These cards were presented to the participants and they were given the instruction to use an answer form. All subjects were tested individually in the presence of at least one of the experimenters.

The first part of the experiment consists of calibration exercises comprising two blocks with 15 trials each. In each block, a reference number and a reference line length are presented to the subjects. In the one block, subjects are instructed to give a numerical estimation (NE) for the given line lengths. In the other block, subjects are required to make a line length production (LLP) for the numbers shown.

After these calibration exercises, the actual experiment starts. We decided to divide the subjects into two groups (12 participants each). In the one group, participants can only use positive integers in the NE condition. In the other group both positive and negative integers are allowed. Except for this manipulation, the experiments are identical for both groups. Both groups also assessed the stimuli with LLP.

The stimuli of our experiment consist of the verbal qualifiers of five commonly used itemized rating scales, the dimensions being: 'evaluation', 'strength', 'agreement', 'satisfaction' and 'frequency' (Figure 4).

Evaluation	Very bad – Bad – Average – Good – Very Good *Reference: reasonably bad = 40 or line length of 20mm*
Strength	Very weak – Weak – Average – Strong – Very strong *Reference: rather strong = 60 or line length of 30mm*
Agreement	Fully Disagree – Rather disagree – No opinion – Rather agree – Fully agree *Reference: agree = 80 or line length of 40mm*
Satisfaction	Very dissatisfied – Dissatisfied – Neutral – Satisfied – Very Satisfied *Reference: rather satisfied = 70 or line length of 35mm*
Frequency	Never – Seldom – Sometimes – Often – Always *Reference: now and then = 50 or line length of 25mm*

Figure 4: Five researched itemized rating scale dimensions, verbal qualifiers and reference stimuli

The stimuli were presented in random order to prevent sequence effects. All verbal qualifiers were evaluated once with a NE and once with LLP.

Results and discussion

Our major concern was to retrieve the two different response strategies: the null point left (NPL) strategy and the middle null point (NPM) strategy. Looking at the data, we immediately see that there is very little support for the NPM strategy. Apparently our subjects have opted for the NPL strategy, but there are also many ambiguous response strategies.

Let r_{ij} denote the magnitude of the assessment (either NE or LLP) of label *j* by participant *i*. And, let a 'strategy' be a sequence of assessments for the 5 successive labels in a scale, where the magnitudes of the assessment of two successive labels is compared. We can now define a 'strict NPL strategy by participant *i*' as "$r_{i1} < r_{i2} < r_{i3} < r_{i4} < r_{i5}$". An 'NMP strategy by participant *i*' can be defined as "$r_{i1} > r_{i2} > r_{i3} < r_{i4} < r_{i5}$". Strategies that are neither strict NPL, nor NMP are classified 'ambiguous'.
This way, the units of analysis consist of the strategy used for one modality by one participant. This way, there are 240 (5 dimensions × 2 modalities × 24 participants) units of analysis.

The NPL strategy was observed 165 times (69%), whereas the NPM strategy occurred only once. The remaining 74 strategies were ambiguous, probably due to the fact that each verbal qualifier was assessed only once, thereby enhancing the impact of inaccurate responses. Noteworthy is the following double null-point strategy, appearing 22 times: "$r_{i1} < r_{i2} > r_{i3} < r_{i4} < r_{i5}$", where r_{i1} and r_{i3} are near zero. This strategy is most salient on the agreement scale (9 times).

Approximately two weeks after testing, we questioned 16 of the 24 participants briefly. We explained the idea of the two different answer strategies and asked if they had ever considered using the NPM strategy. Three participants thought of this strategy as a possible way of answering. We also tried to find out which scales would be more prone to either strategy, by presenting the list with the 5 dimensions and asking on which scale(s) they would most/least likely use an NPM strategy (even though they preferred an NPL strategy)[i]. Results are summarised in table 1.

Table 1: Overview of the responses on the follow-up questions

Most NPM		Least NPM	
agreement	12	frequency	11
satisfaction	6	evaluation	4
strength	2	strength	4
none	1	satisfaction	3

This research offers little support for the two different strategies as found by Hofmans et al (2004), particularly for the NPM strategy. Only one participant employed the NPM strategy for one dimension and only in the LLP condition. There were 30% ambiguous strategies, meaning they were neither strictly NPL nor NPM.

However, 22 of the 74 ambiguous strategies are what we could call "double null-point strategies" which occur most for the agreement scale. A possible explanation is that these participants use an NPL strategy, but consider a label like "no opinion" to belong to some other dimension, and in doing so, create a hybrid strategy of NPL and NPM.

On counting the strict NPL strategies, the frequency scale seems to be the most prone to this strategy (41/48). This is also supported by the follow-up questions: 11 of the 16 participants mentioned the frequency scale as being one of the least NPM-prone scales. Agreement, followed by satisfaction is, according to the participants, most likely scale to elicit NPM strategies. We believe frequency is more unanimously seen as being a one-dimensional scale and by that increasing the chances on a strict NPL strategy and decreasing the chances for (close to) NPM strategies.

In our experiment, some participants could use only positive integers for their assessments, while others could also use negative integers. Only 2 of the 12 participants in the last condition used a negative number, and both did so only once. Why is it that so few negative numbers were used? To us, it seems possible that participants might use these negative numbers more often if they were not asked to draw line lengths as well. Since line lengths cannot be drawn smaller than zero, the NE answers are possibly affected by the LLP answers. One participant mentioned in the follow-up questions that she would have used negative numbers if given the opportunity (by placing a zero on the middle label).

General discussion

Very mixed results were obtained in the three experiments discussed above. While two experiments showed evidence for the existence of two response strategies, NPM and NPL, one experiment, especially designed to replicate these strategies, failed to reproduce them. The caution with this third experiment is that the group of respondents was rather small, increasing the possibility that coincidentally nobody with an NPM strategy was selected. Another alternative explanation is that the strategies, especially the NPM strategy, were an artefact of the experiments and do not exist outside of these situations. If however these strategies are real, one can draw a number of conclusions from our experiments.

A first conclusion is that the two response strategies, NPM and NPL, depend on the perception of the scale by the respondent. Not every respondent perceives the same scale in the same way as the strategies are very subjective. A second conclusion is that not all dimensions are susceptible to these two strategies. The two dimensions most vulnerable for the NPM strategy are apparently agreement and satisfaction. A third conclusion is that for some people the instructions in a cross-modality matching experiment can be suggestive. Since with most modalities no negative values can be allocated, people are pushed towards a NPL strategy. When we cancel this limitation, some participants shift from an NPL strategy to an NPM strategy. Problematic for category-scaling experiments is the failure of most modalities since they do not allow negative appraisals. For this kind of experiments the only valuable modality is numeric estimation, but using this modality alone does not allow us to validate the responses of the participants. Perhaps the cross modality paradigm, as used in these experiments, is not appropriate for experiments with category rating scales.

References

Breakwell, G. M., Hammond, S. & Fife-Shaw, C. (2000). Research methods in Psychology. Sage Publications, Londen.

Cools, W., Hofmans, J., Baekelandt, S., Theuns P. (2004). The numeric estimation of verbal qualifiers measuring level of agreement: differences in response patterns. Paper presented at Fechnerday 2004, The 20th Annual Meeting of the "International Society for Psychophysics", Coimbra, Portugal.

Hofmans, J., Baekelandt, S., Cools, W., Theuns, P. (2004). *The influence of reversing the rating scale on the intensity perception of the anchors and on the response pattern.* The 20th Annual Meeting of the "International Society for Psychophysics", Coimbra, Portugal.

Lodge, M. (1981). *Magnitude scaling: quantitative measurement of opinions*, Sage University Paper series on quantitative applications in social science. Series no. 25, Sage publications: London.

Poulton, E.C. (1989). *Bias in Quantifying Judgments,* Lawrence Erlbaum Associates Publishers: Hove and London.

Stevens, S. S. (1966). *A metric for social consensus: methods of sensory psychophysics have been used to gauge the intensity of opinions and attitudes.* Science, 151, 530-541.

---

[i] The respondents were allowed to select more than one scale/dimension. For example, some participants mentioned both agreement and satisfaction as the dimensions/scales on which they would 'most likely use NPM', while goodness and strength were often mentioned together when asking for the dimensions/scales on which they would 'least likely use an NPM strategy'.

SEMANTIC AND PERCEPTUAL REPRESENTATIONS OF COLOR

Kimberly A. Jameson
Institute for Mathematical Behavioral Sciences
University of California at Irvine, Social Science Plaza, Irvine, CA 92697-5100
e-mail: kjameson@uci.edu. http://aris.ss.uci.edu/∼kjameson/kjameson.html

Abstract

Perceptual psychology widely operationalizes *color appearance* as a construct with very close, even isomorphic, ties to color naming structure. Indeed, a considerable body of psychological and psychophysics research uses naming–based tasks to derive structural properties of color appearance space. New research investigating the relations linking color similarity and color naming structures suggest that assumptions involving strong structural correspondences between appearance and naming are unfounded. Such research also reveals *(i)* features of the phenomena for which cognitive and perceptual learning processes play significant roles in establishing individual's color naming similarity structures and *(ii)* features of the mechanisms underlying stable color naming systems and the ways such shared systems relate to varying individual perceptual color experience. Empirical support for these is summarized, as are suggestions for exploring the largely uninvestigated cognitive processes underlying color appearance and naming similarity relations.

Much research on the cognitive processing of color appearance assumes that for all trichromat observers well–specified perceptual relations exist for predicting how color names will be used to label and partition color appearance space. This assumption is present in theoretical descriptions of shared color lexicons, in the practiced aggregation of color naming data across individuals to derive color naming norms, and in empirical tasks that use lexical labels when assessing individual perceptual salience and individual color space similarity relations. Actually, the perceptual basis for color naming is far from normatively uniform. Simple anomalous trichromats – who in the real world behave as normal trichromats – can have extremely different color equivalence classes compared to normal trichromats, with both smaller and larger metameric discrimination ellipses compared to normal observers (Regan et al. 1994). Even among normal trichromats, observers can experience rather deviant (i.e., $\sim$3 s.d.) red–green color weakness (National Reseach Council 1981), suggesting that discrimination differences among *normals* are enormous (Kaiser & Boynton 1996, 343). Large differences among trichromats are also found for "fundamental" unique hue settings, and no uniformity exists in the perceptual distances of unique hue ranges (Kuehni 2001, 2004, 2005). Such results do not support the view that uniform perceptual processing is the basis for color naming and categorization findings (Kay & Regier 2003).

The absence of a uniform perceptual basis for *normal* appearance and naming representation raises some interesting questions. For example, *what is the basis for the cognitive representation of color appearance, and for color categorization and naming systems?*, and *what psychological processes play a role in the representation and maintenance of such systems*? Here it is shown that consideration of the constraints on color representation under normal perceptual variation can provide new insights into the cognitive mechanisms contributing to individual color appearance and naming cognitive representation.

Trichromat Observer Variation

To clarify the *basis of individual color appearance and naming representation*, it seems uncontroversial to suggest that cognitive representation of color appearance similarity must vary in a manner that accords with an individual's variant of color perception. Figure 1 depicts color relations for two trichromat observers. Both panels (a) and (b) show a triangular slice of uniform *lightness* from a

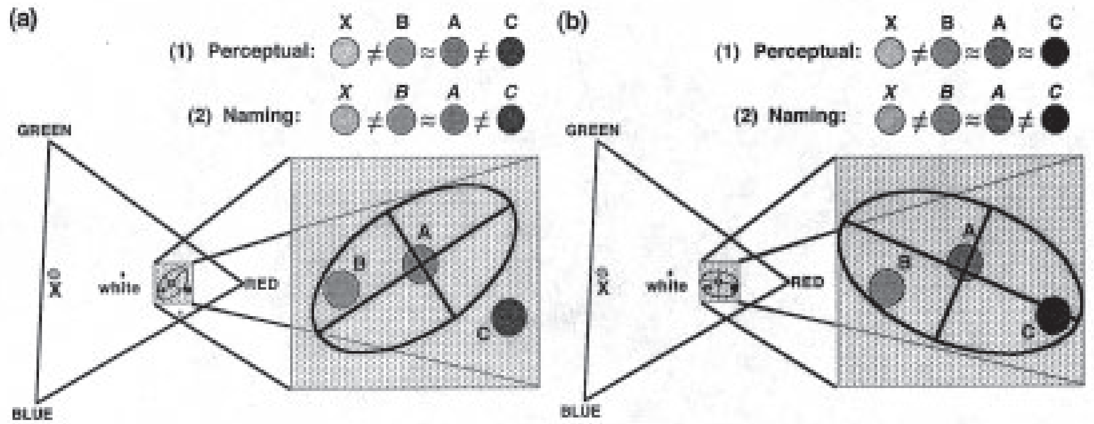


Figure 1. Schematic of a hypothetical *normal* trichromat color equivalence region showing congruent (1) perceptual and (2) naming relations (panel (a) at left). Similarly, panel (b) depicts hypothetical color equivalences for an *anomalous* trichromat. Depicted ellipses are not empirically determined and are not drawn to scale. Analogous differences in trichromat metameric class ellipses were empirically identified by Regan et al. (1994).

three-dimensional color appearance solid consisting of vertices Red, Green and Blue – analogous to a *CIE* chromaticity diagram. Within the triangular color plane a miniaturized region of color space is shown, and is also enlarged at right. Note in the miniaturized view at left, the stippled area of the space shows three *reddish* color samples (labeled A, B, C in the enlarged stippled view), and a more distant area shows a *bluish-green* sample labeled X. Each Figure 1 panel presents two expressions describing different observer's perceptual and naming relations for A, B, C and X.

Figure 1, panel (a) illustrates that *normal trichromat* perceptual similarity relations can accord with color naming relations. That is, perceptually similar appearances *A* and *B* are both distinguished from *C*, and are named cogruently (as relations 1 and 2 suggest). For example, appearances *A* and *B* may both be named *crimson* and *C* named *maroon*. By comparison, Figure 1(b) shows that color appearance similarity relations of an *anomalous trichromat* (a trichromat with one or more shifted retinal photopigments causing systematic differences in the observer's perceptual equivalence classes) can (*i*) differ from that of the normal trichromat in Figure 1(a), but can also (*ii*) be named congruent with the "normal" color naming similarity relations in Figure 1(a). Thus, perceptual and naming relations of an anomalous trichromat can be incongruent as shown in panel (b), while those for a "normal" trichromat are congruent. The separate appearance and naming representations illustrated in Figure 1 are suggested as linked by different cognitive *color–naming functions* dependent on observer type (Jameson & Alvarado 2003a).

Figure 1's illustration that color naming can be congruent across observers when color appearance relations are not, implies an idea central to the present theory. Namely, while color appearance similarity can vary, shared color naming similarity relations are normatively stable across a variety of observer types due to socio–cultural and pragmatic constraints present in individual color communications (Jameson 2005). Figure 1 also exemplifies that perceptual variations due to inherited color perception abilities need not impact the sharing of linguistic color relations within an ethnolinguistic group, and suggests that color categorization and naming universals within and across ethnolinguistic groups are not attributable to shared privileged perceptual salience across individuals (Jameson & Alvarado 2003b, Jameson 2005a, 2005b). Instead, individuals share a lexicon's relational structure by communicating with members in their society in ways that reinforce and maintain the stable communication code. The clear purpose of the cognitive *color–naming function* in this scenario is to strive for maintanence of a shared naming–system equilibrium despite individual variation in perceptual representation or other naming idiosyncracies. Thus, similar to a dichromat, the anomalous trichromat may not perceptually distinguish some colors samples, but she still possesses the *shared naming relations* of normal trichromats (cf., Shepard & Cooper 1992).

The relational mappings of appearance to lexical categories achieved by an individual's color–naming function is not uniform across color space, or context. Trichromats' perceptual and linguistic relations can be identical for large color differences, but they need not be identical for smaller color differences. In dichromats, large color differences may be undetectable perceptually but present linguistically. The existence of shared lexical representations, distinct from color perception representations, partially explains why dichromat observers can be undetected in everyday social interactions with trichromats, and suggests a highly cognitive *automatic*) meta-awareness about one's own color experience compared to others in the culture. Dichromats understand that Trichromats perceive *red* and *green* as opposing categories, and in everyday interactions they are only at a disadvantage for naming when they have no other cues except color properties to help differentiate two items within their confusion classes.

Retinal Tetrachromat Observers

To clarify *psychological processes that play a role in the development and maintenance of color naming and appearance representations*, it is useful to examine cognitive processing across additional observer types. Recent research has show that some observers possess the genetic potential for more than the normal numbers of retinal cone classes, and that such observers can experience forms of *weak* or *strong* tetrachromacy (Jordan and Mollon 1993). It is known that color discriminability and dimensionality vary as a function of observer retinal phenotype. The number of just-noticeably-different color perceptions experienced by rod monochromats consists of 10^2 different color experiences (all black and white combinations), dichromat individuals experience 10^3 different color experiences, trichromats approximately 10^6 different color experiences. The obvious question is whether a retinal tetrachromat experiences analogous differences. Retinal tetrachromats occur due to additional variants of photopigment opsin genes acquired by X-chromosome inheritance. When allelic variations for M– and L–cone classes exist at certain positions on the genetic array, shifts in spectral response sensitivity occur that impact color perception. The range and variety of photopigment variants is surprising and implies that, in some populations, retinal processing is almost certainly more varied than originally anticipated by color theory. Although the actual phenotype frequency is uncertain, it is known that a considerable percentage of Caucasian females have the *genetic potential* to express four classes of retinal photopigments (Sharpe *et al.* 1999). Recent empirical studies suggests that retinal tetrachromat genotypes correlate with differences in color categorization, naming and similarity (discussed below), thus presenting further opportunity to analyze the relations between individual color experience and shared color naming representations.

Figure 2 illustrates hypothesized perceptual and linguistic relations for a retinal tetrachromat, depicting, for comparison with Figure 1, a perceptual equivalence–class region potentially available to a retinal tetrachromat. In Figure 2 a small region of *reddish* color appearance is enlarged at right where a hypothetical tetrachromat equivalence–class is centered on sample A and differs from that shown in Figure 1, suggesting shifted or compressed equivalence–class contours compared to a trichromat (generalized from Jameson *et al.* 2001). These hypothetical retinal tetrachromat *perceptual relations* imply that in contrast to Figure 1 observers, a retinal tetrachromat may perceptually distinguish between sample A and B. Despite this perceptual difference, and similar to Figure 1's anomalous trichromat, the retinal tetrachromat uses a shared color lexicon influenced by the society's trichromat majority (i.e., expression (2)). This suggests that a retinal tetrachromat may be capable of both encoding greater lexical specificity than that found in a normative color lexicon, and greater numbers of categorical distinctions than are agreed upon by trichromats in their society. Of course, the trichromat also perceptually re-

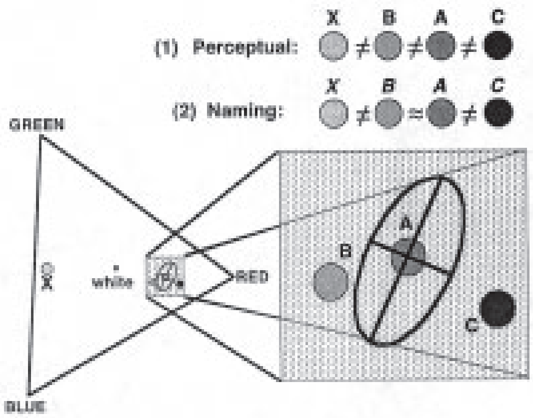


Figure 2. Schematic of a color space region for a hypothetical retinal tetrachromat with differing (1) perceptual and (2) naming relations. Equivalence–class ellipses are drawn strictly for illustrative purposes.

solves more color distinctions than are represented by the lexicon (compared to a Dichromat with *less* perceptual specificity). However, while a society's color lexicon may be adequate as a trichromat color communication code, it may be inadequate, or lack sufficient specificity, for tetrachromat observers.

Earlier it was suggested that social influences and perceptual learning help smooth out color–naming discord potentially arising from perceptual differences among members of a culture. An essential component in this is a tendency for *linguistic charity*, or flexible discourse, among members of a society (Jameson & Alvarado 2003a). For example, as with dichromats, tetrachromat observers may learn to accept and comfortably use a comparatively imprecise mapping of color appearances to color language and categories. A retinal tetrachromat child developing in the company of trichromats may learn color categories primarily by discovering that groups of objects that appear different in color to the child are consider as color–matched by other people. After reliable exposure to these learning experiences such an observer could develop a personal definition of color similarity that says: "Color *matching* denotes when two things have *almost the same* color appearance to me, although other people report seeing them as identical." In this example, the retinal tetrachromat's cognitive construct of a color–match differs from a trichromat's. The net result is that potential disagreements of color labeling among individuals with varying perceptual abilities are minimized. Such a naming function may play a role in other individually varying representations involving color compatibility, color preference, color memory. How tetrachromat perceptual relations might differ from linguistic relations, and how they vary from a trichromat norm, is clarified by recent results on color processing behavior which are now described.

Color Processing Behaviors Correlated with Potential Retinal Tetrachromacy

Psychophysical discrimination paradigms have produced evidence of weak Tetrachromacy (Nagy et al. 1981, Jordan & Mollon 1993), but only one case of reliably strong tetrachromacy (Jordan & Mollon 1993). There are good reasons why some psychophysical viewing circumstances might not register tetrachromat perceptual differences (see Jameson *et al.* 2001). In contrast, empirical assessment under naturalistic viewing circumstances, using a variety of cognitive judgments (color similarity, color categorization, and color naming), show reliable correlations between retinal tetrachromat genotypes and differences in color behavior.

Jameson *et al.* (2001) found that females with retinal tetrachromat genotypes experience substantial

Subject Partition	*M*	*SD*	n
(1) Females with retinal tetrachromat genotypes (opsin gene heterozygotes)	10.0	2.96	23
(2) Trichromat Females	7.6	1.80	15
(3) Trichromats (Females and Males)	7.3	1.93	37
(4) Dichromat Males	5.3	1.53	4

Table 1: Means of Individual Median Spectral Delineations for Four Subject Partitions.

differences in color perception compared to normal female trichromat controls. They used a task in which subjects delineated categories in a diffracted spectrum subjectively appearing as a luminous "rainbow," and hypothesized that the ability to perceive and delineate chromatic bands in the spectrum was a function of perceiving noticeable differences in spectral wavelengths. Such differences were expected to covary with the number of retinal photopigment classes possessed. Their results showed significant covariation of tetrachromat genotypes with increased spectral delineation behavior.

Table 1 shows that the spectral bands a subject delineates systematically varies with the number of photopigments a subject is presumed to express (Jameson *et al.* 2001). HERE Classification of subject partition (1) is inferred strictly from the genotype analysis determining heterozygote and is probabilistically linked to the four-photopigment phenotype (with an estimated 56% incidence of genotype occurrence). Partitions (2) to (4) are based on results from both genotype tests and color-vision screening tests. Partition (2) is a sub-partition of group (3). As expected, dichromat individuals delineate fewer chromatic bands than trichromats (Student's t-test, two-tailed, equal variance $p<.05$). Male Trichromats were not significantly different from female Trichromats ($p=.44$). And a significant difference ($p<.01$) was found between female retinal tetrachromat genotypes (or *heterozygotes*) and trichromats (male and female) subjects. However, the most stringent test rules out possible gender differences in socialization: The number of bands observed between the two female groups (rows 1 and 2 of Table 1) is significantly different ($p<.01$). Overall, Table 1 indicates a systematic relationship between the observed number of bands delineated by subjects and the number of photopigments they are presumed to express. These results suggest that color experience for retinal tetrachromat females is complex compared to "normal" trichromatic color vision; or, less conservatively, that some females show signs of tetrachromacy. (Although whether it is a weak or strong tetrachromacy is unknown.)

Similarly, Jameson, Bimler & Wasserman (2005) found support for a tetrachromat perceptual difference by comparing standardized color vision assessment results between retinal tetrachromat genotypes and three–gene trichromatic genotypes. Novel multidimensional scaling analyses revealed that the Farnsworth–Munsell 100 Hue Test, identifies some retinal tetrachromat individuals (who otherwise exhibit above-average color discrimination) with a non-normative diagnosis, suggesting that such tests do not appropriately capture a tetrachromat's non-deficient perceptual variation. Figure 3 shows that 32% of genetically identified heterozyotes were diagnosed as false–positive deficient on the F–M 100 test (by Z-values at least 1 s.d. from normal) – all of which had otherwise normal color vision – and among these were heterozygotes who perceived the greatest number of spectral delineations (seen in Figure 3's top partition). When the F–M 100 test is used as a screen to eliminate subjects with color perception defects (as is the common practice), such misclassifications would result in the omission of some non–defective retinal tetrachromats from the "normal" subject sample tested.

Sayim, Jameson, Alvarado & Szeszel (2005) assess cognitive color behaviors of retinal tetrachromats using triad similarity for color samples and color names across different stimulus sets, and find that some measures differentiate retinal tetrachromat genotypes from trichromatic genotype controls. Analyzing females separately from males, they find that measures of group agreement and consistency increase with opsin genotype complexity. L–cone dimorphisms seem instrumental in the behavioral

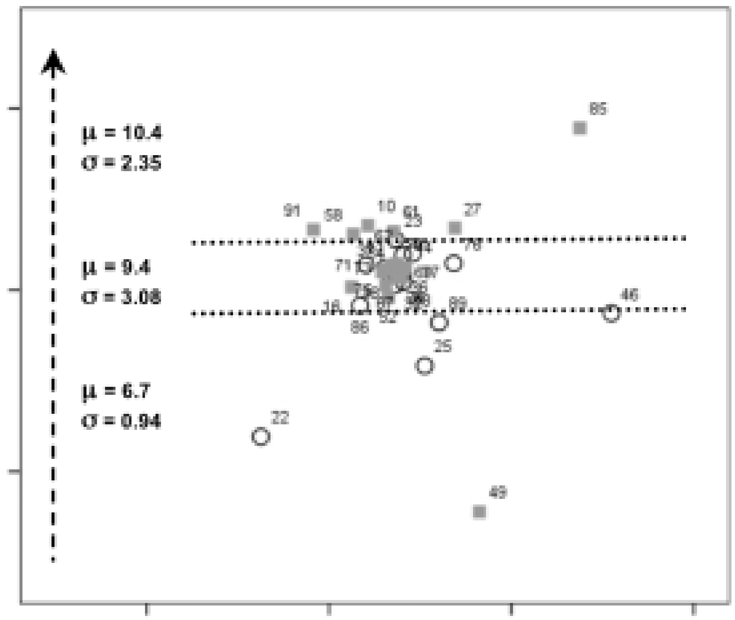


Figure 3. Two-dimensional dissimilarity scaling of F–M100 performance for 37 female subjects. Two horizontal lines were drawn *ad hoc* to emphasize the association between F–M100 performance, genotype, and color perception in the Jameson *et al.* (2001) spectral delineation task. Square gray symbols denote heterozygote females and open circles denote homozygous females. The vertical dashed arrow is a rotated regression line along which banding behavior increases and decreases. Mean and standard deviation banding for partitions are shown.

differences because from among all subject groups examined, only the L-opsin gene heterozygotes exceeded criterion on all measures evaluated. Sayim *et al.*'s (2005) results are strong evidence supporting the distinct perceptual and naming representations described above.

Figure 4 summarizes consensus theory (Batchelder & Romney 1988) results for four female genotype groups (Sayim *et al.* 2005). Consensus scores for color triads (left panel) and word triads (right panel) show that between groups significant differences are not seen for the perceptual color triads, although significant differences are found for word triads. Specifically, word triad consensus tends to increase with increases in opsin genotype complexity. Of particular interest is the group with only L-opsin dimorphisms (n=6) which was unique in showing differences in *(i)* consensus across global and local **color triads** tested, and *(ii)* good levels of consensus in local **naming triads**, for which all other subjects showed low consensus. A speculative interpretation of *(i)* is that perceptual variation occurring within the *only L-opsin dimorphism* group is not systematically modeled by the three dimensional color constraints imposed of our CRT–based stimuli. In turn, for this group alone, this could lead to ambiguity in color similarity in the **color** triads tested, producing differences in consensus scores. Although L-cone variation may alter red CRT phosphor sensitivity by 7% (Golz & MacLeod 2003), we have yet to conduct the detailed analyses needed to identify the causes underlying the triad similarity variation we observed. Next, a speculative interpretation of *(ii)* is that subtle differences in perception arising from expressed dimorphisms might bias such females towards developing color expertise by cognitively heightening color awareness relative to females without such dimorphisms. Over a life–time this subtle increase in color awareness might lead to the cultivation of color naming expertise, possibly producing a more robust lexical code and greater naming consensus as seen in the word triad results for the *only L-opsin dimorphism* group. While both speculative interpretations support the suggestion that observers from the *only L-opsin dimorphism* group may experience something different from "normal" trichomacy, further psychophysical research is needed to determine the nature of the perceptual variation and its influences on color representation. Some have suggested a *dimensional* color differ-

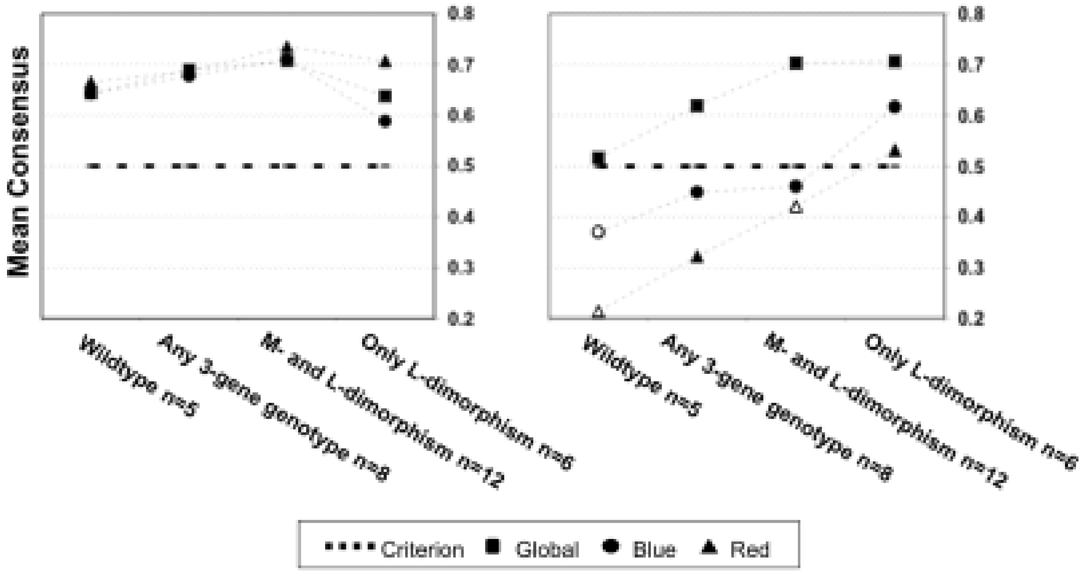


Figure 4. Mean Consensus Scores for Color triads (left panel) and Word triads (right panel) for four Female Genotype Groups. Open symbols denote analyses with low mean consensus and eigenvalues failing to show a single dominant factor. Global results are based on a selected set of color stimuli that span categories (*e.g.,* purple, green, blue, red, *etc.*), whereas Blue and Red results are "local" results based on separate sets of blue and red color stimuli varying strictly within-category. See Sayim *et al.* 2005 for details.

ence may exist in retinal tetrachromat observers, this however would require a re–assessment of the widely received assumption of *neural trivariance*, thus it seems prudent to further investigate these phenomena before asserting that dimensional differences are found perceptually.

Summary

Interesting cognitive implications arise by separating color appearance and naming representations. Jameson (2005b) provides an extensive description of these. Clearly, analyses of different observer types are useful for understanding different cognitive influences on color representation. Psychophysical theories of color representation should at a minimum allow for the possibility that some observers experience relational or dimensional differences in color appearance, and still consistently share a normative naming similarity structure. The separate perceptual and semantic representations suggested are linked by a flexible mapping function, or a *cognitive color–naming function* (Jameson & Alvarado 2003a), which should robustly map large color differences to color categories for both normal and anomalous trichromats, but it should be comparatively less robust for mapping small color differences (especially near boundaries). As suggested earlier, different observer types, dichromats, trichromats, anomalous trichromats and possibly functional tetrachromats, could all acquire and use (with varying efficiency) a culturally normative naming system, but they may learn such a system by different manners using different strategies. Jameson and Hurvich (1978) suggested that dichromats learn to use and recognize normative hue terms mappings through correlation with brightness despite the inability to differentiate some hues. By comparison, tetrachromats may implicitly learn that the normative trichromat category tolerance permits a wider range of perceptual variants than would be distinguished in a tetrachromat category structure. It is not necessary to accept the existence of human tetrachomacy to recognize that some of the psychological processing factors described here dictate the separation of individual color appearance and naming representations, or to accept that the cognitive issues separating such representations are greatly under researched in the extensive interdisciplinary color representation literature.

References

Batchelder, W. H. & Romney, A. K. (1988). Test Theory without an Answer Key. *Psychometrika, 53*, 71–92.

Golz, J. & Macleod, D. I. A. (2003). Coloimetry for CRT displays. *Journal of the Optical Society of America, A 20*, 769–781.

Jameson, D. & Hurvich, L. M. (1978). Dichromat color language: “Reds” and “Greens” don’t look alike but their colors do. *Sensory Processes, 2,* 146–155.

Jameson, K. A., Bimler, D. & Wasserman, L. (2005). Re–Assessing Perceptual Diagnostics for Observers with Diverse Retinal Photopigment Genotypes. In *Progress in Color Studies 2: Cognition*, Pitchford, N. J. & Biggam, C. P. (Ed.s). Amsterdam: John Benjamins Publishing Co. In–press.

Jameson, K. A. (2005a). Why *GRUE*? An Interpoint–Distance Model analysis of Composite Color Categories. *Cross–Cultural Research, 39(2)*, 159–194.

Jameson, K. A. (2005b). Culture and Cognition: What is Universal about the Representation of Color Experience? *Journal of Cognition & Culture, 5(3/4)*, 29–85.

Jameson, K. A. & Alvarado, N. (2003a). The Relational Correspondence between Category Exemplars and Naming. *Philosophical Psychology, 16(1)*, 26–49.

Jameson, K.A. & Alvarado, N.. (2003b). Differences in Color Naming and Color Salience in Vietnamese and English. *COLOR Research and Application, 28(2)*, 113–138.

Jameson, K. A., Highnote, S. & Wasserman, L. (2001). Richer color experience for Observers with multiple photopigment opsin genes. *Psychonomic Bulletin & Review, 8(2),* 244–261.

Jordan, G. & Mollon, J. D. (1993). A study of Women Heterozygous for Colour Deficiencies. *Vision Research, 33*, 14951508.

Kaiser, P. K. & Boynton, R. M. (1996). *Human Color Vision, 2nd Edition*. Optical Society of America, Washington, DC.

Kay, P. & Regier, T. (2003). Resolving the Question of Color Naming Universals. *Proceedings of the National Academy of Science, 100*, 9085–9089.

Kuehni, R. G. (2001). Determination of unique hues using Munsell color chips. *COLOR Research and Application, 26,* 61–66.

Kuehni, R. G. (2004). Variability in Unique Hue Selection: A Surprising Phenomenon. *COLOR Research & Application, 29*, 158–162.

Kuehni, R. G. (2005b). Focal Variability and Unique Hue Stimulus Variability. *Journal of Cognition & Culture, 5(3/4)*.

Nagy, A. L., MacLeod, D. I. A., Heyneman, N. E., & Eisner, A. (1981). Four cone pigments in women heterozygous for color deficiency. *Journal of the Optical Society of America, A, 71*, 719–722.

National Research Council (1981). *Procedures for Testing Color Vision: Report of Working Group 41.* Committee on Vision Assembly of Behavioral and Social Sciences. National Academy Press, Washington, D.C.

Regan, B. C., Reffin, J. P., & Mollon, J. D. (1994). Luminance noise and the rapid detemination of discrimination ellipses in colour deficiency. *Vision Research, 34,* 1279-1299.

Sayim, B., Jameson, K. A., Alvarado, N. & Szeszel, M. K. (2005). Semantic and Perceptual Representations of Color: Evidence of a Shared Color–Naming Function. *The Journal of Cognition & Culture, 5(3/4)*,165–220.

Sharpe, L. T., Stockman, A., Jägle, H. & Nathans, J. (1999). Opsin Genes, Cone Photopigments, Color Vision, and Color Blindness. In *Color Vision: From Genes to Pereption.* Gegenfurtner, K.& Sharpe, L. T. (Ed.s). (p. 3–52).

Shepard, R. & Cooper, L. (1992). Representation of colors in the blind, color–blind, and normally sighted. *Psychological Science, 3,* 97–103.

Global vs. Local Perception in High-Functioning Autism: Architecture, Capacity, and Decision-stopping Rules

Shannon Johnson, James Townsend, Leslie Blaha, Robin Murphy, and Julie Stout
Department of Psychology, Indiana University, Bloomington, IN 47405

Abstract

Accumulating evidence from studies of global-local processing suggests a lack of global advantage and lack of global interference in autism spectrum disorders (ASD). Despite the potential importance of differences in this fundamental aspect of information processing, the reason(s) for deviations in global-local performance remains unclear. To determine the cognitive strategies underlying global-local performance in individuals with and without an ASD, we applied mathematical models of information processing. Six individuals with Asperger's Disorder, 8 age-matched comparison subjects, and 13 adult controls completed selective and divided attention global-local tasks. Specifically, the double factorial paradigm (DFP) and capacity measures developed by Townsend and colleagues (Townsend & Nozawa, 1995; Wenger & Townsend, 2000) were used to determine the processing architecture (parallel or serial), stopping rule (self-terminating or exhaustive), and efficiency (limited, unlimited, or super capacity) for each participant. Given the heterogeneity of cognitive abilities in ASD, analyses were completed at the individual and group level. Overall, there was no evidence of a local bias in the ASD group. In contrast to previous findings, both global and local interference effects were consistently demonstrated by participants in all groups, including all ASD subjects. Finally, modeling results indicated that cognitive strategies employed during the global-local tasks varied across individuals and that some ASD participants used approaches that were distinct from control subjects. Results suggest that specific cognitive processing strategies may explain previous global-local performance differences in ASD.

Although social deficits define autism spectrum disorders (ASD), differences in perceptual and cognitive processes are also frequently reported. In particular, several studies have demonstrated differences in global-local processing in ASD. Initial studies (Jolliffe & Baron-Cohen, 1997; Shah & Frith, 1983), using clinical measures, demonstrated fast and accurate processing of local details in ASD relative to control subjects. It was argued that a local processing bias exists in ASD, and that global processing is impaired. More recently, several studies using traditional cognitive paradigms (e.g., Navon, 1977) have reported that ASD groups demonstrate a lack of global precedence and lack of global interference (Mottron, Peretz, & Menard, 2000; Plaisted, Swettenham, & Rees, 1999; Rinehart, Bradshaw, Simon, Brereton, & Tonge, 2000); both findings are in contrast to studies of normal cognition. However, these studies have not supported previous claims that global processing is impaired. The purpose of the current study was to further define the nature of global-local processing in ASD using both selective and divided attention tasks. In particular, we investigated the utility of new stimuli developed to accommodate the double factorial paradigm (DFP; Townsend & Nozawa, 1995). By utilizing the mathematical modeling approach, we are able to determine whether information was processed in a serial, parallel, or coactive manner (i.e., what was the processing architecture?); and whether the information was processed exhaustively or if processing was self-terminated when sufficient information was acquired (i.e., what type of stopping rule was employed?). Finally, given the known heterogeneity of cognitive abilities is ASD, we approached these questions at an individual and group level.

Method

Participants

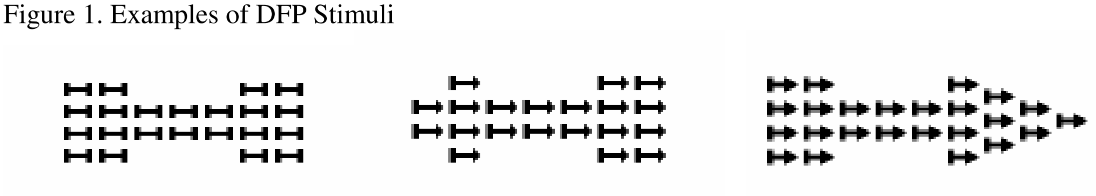
Six individuals diagnosed with Asperger's Disorder (ASP), 8 age- and IQ-matched normal comparison participants (AGEC), and 13 adult controls (ADULT) completed selective and divided attention global-local tasks. Specifically, the double factorial paradigm (DFP) and capacity measures developed by Townsend and colleagues (Townsend & Nozawa, 1995; Wenger & Townsend, 2000) were used to determine the processing architecture (parallel or serial), stopping rule (self-terminating or exhaustive), and efficiency (limited, unlimited, or super capacity) for each participant. Participant IQ was estimated using the Vocabulary and Block Design subtests of the Wechsler Abbreviated Scale of Intelligence (WASI). The groups were closely matched with regard to estimated IQ, with means in the average to above average range of intellectual functioning.

Tasks and Procedures

For each subject, all tasks were administered on the same computer, using DMDX software. Each stimulus was presented for 65ms and response time was measured from stimulus offset. A mouse was used as the input device and was centered in front of the subject on the desk. Using the index finger of each hand, participants responded by pressing the appropriate mouse button (right or left) as quickly as possible. The DFP requires two levels of salience for two factors (global & local); the less salient level is referred to as 'low' and the more salient is 'high'. Examples of stimuli are presented in Figure 1.

For all selective attention tasks, participants were instructed to attend to one dimension, global or local, and to respond by pressing the right mouse button if the specified dimension was pointing right or the left button if not pointing to the right (i.e., left or neutral). The non-target dimension or *distractor* was neutral, compatible (i.e., redundant with target dimension) or incompatible (i.e., opposite direction). A total of eight Garner and Stroop-like tasks were administered. These tasks are summarized in Table 1.

Figure 1. Examples of DFP Stimuli



Global and Local = neutral — Global = left, Local = right; Global and local salience = low — Global = right, Local = right; Global and local salience = high

Table 1. Summary of Selective Attention Tasks

Task	Target	Non-Target Distractor	# of Target Trials
1	Right Local	Right Global	160
2	Right Global	Right Local	160
3	Right Local	Left Global	160
4	Right Global	Left Local	160
5	Right Local	Mixed Global (R, L, & N)	180
6	Right Global	Mixed Local (R, L, & N)	180
7	Right Local	Neutral Global	160
8	Right Global	Neutral Local	160

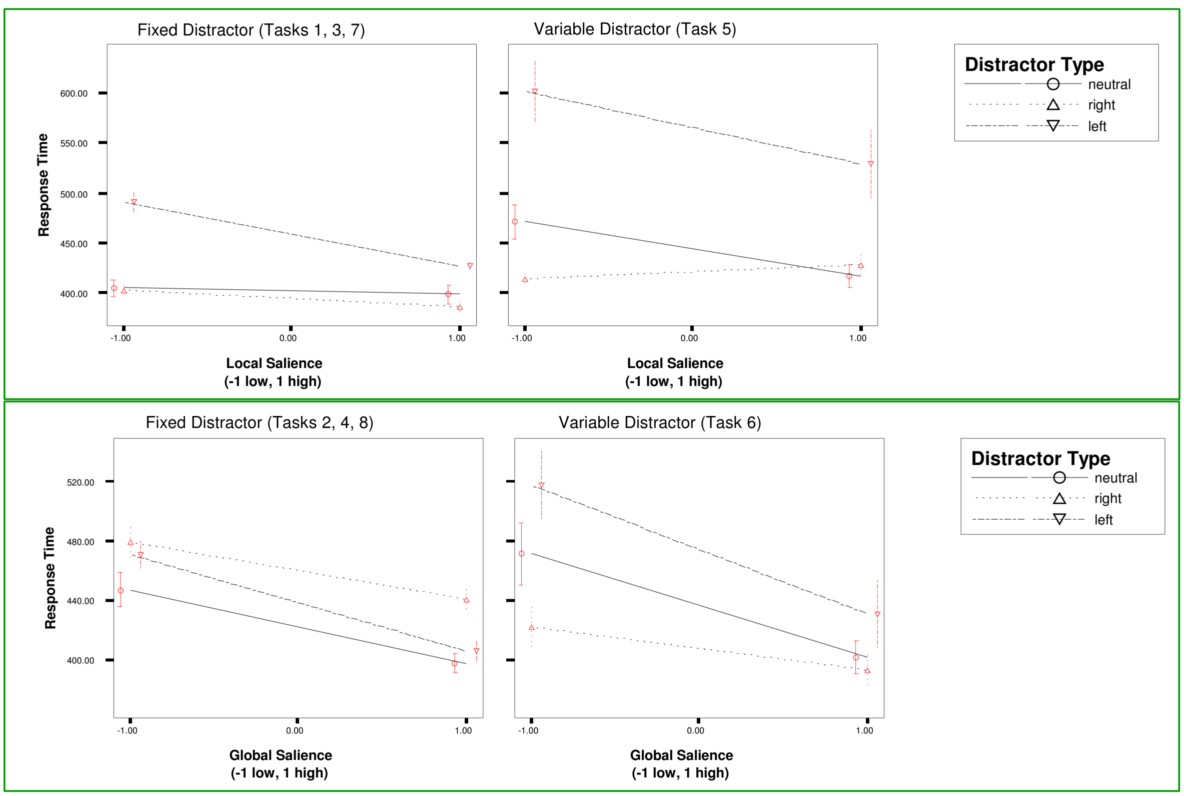
For all divided attention tasks (DFP), participants were instructed to respond by pressing the right button if any arrow pointed to the right or the left button response if no arrows pointed right. The stimuli included neutral, global right/local neutral, local right/global neutral, or global right/local right (redundant case) dimensions. The target dimension was not specified.

Results

Interference Effects

The majority of subjects completing all tasks (19/23) responded significantly faster to global versus local stimuli. There was no evidence of a local bias in the ASP group. In contrast to previous findings, both global and local interference effects were consistently demonstrated by participants in all groups, including all ASP subjects. Comparisons of distractor type (neutral, incompatible, compatible/redundant) indicated that for both local and global stimuli, incompatible distractors slowed response times. This effect was very robust in Tasks 5 and 6, where the distractor type was varied randomly throughout. In addition, similar findings were also present (although typically attenuated) in the single distractor tasks (Task 1-4, 7-8) for several participants. Figure 2 illustrates this finding in two subjects. In contrast to previous findings, the interference effect was consistently demonstrated by members of all groups, including all ASD subjects.

Figure 2. Interference Effects



Redundancy Effect
Within the divided attention task, response times for the single dimension stimuli (i.e., local right/global neutral and global right/local neutral) were compared to response times for redundant items (i.e., both local and global right facing arrows). For all subjects, responses were significantly faster for redundant stimuli compared to both single dimension target-present items.

Salience Manipulations
Across tasks and for all subjects, there was consistent evidence that response times were significantly faster for high salience dimensions relative to low salience dimensions. This result is a prerequisite for proceeding with the cognitive modeling based on the DFP responses.

<u>Cognitive Modeling Results</u>

Architecture and Stopping Rule
The DFP (Townsend & Nozawa, 1995) uses two rigorously defined contrast measures to non-parametrically describe the architecture and stopping rule of a cognitive processing system. Using the factorial combination of salience manipulations in the redundant target trials, the DFP measures are:

Mean Interaction Contrast (MIC):

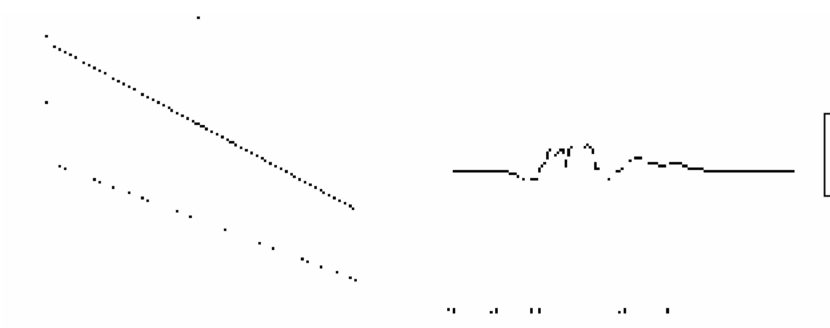
$$MIC = (\overline{RT}_{LL} - \overline{RT}_{LH}) - (\overline{RT}_{HL} - \overline{RT}_{HH})$$

Survivor Function Interaction Contrast (SIC):

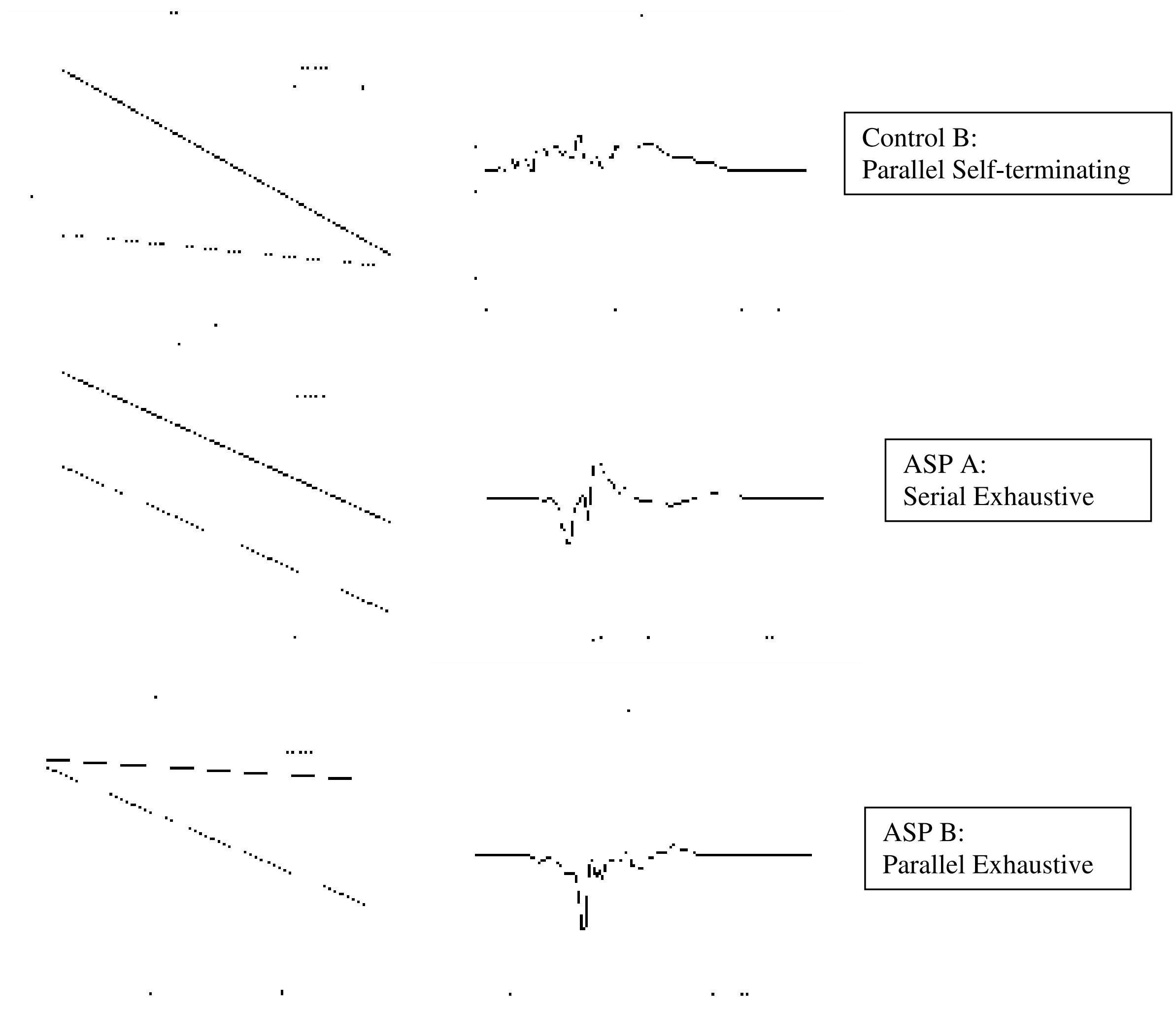
$$SIC(t) = (S(t)_{LL} - S(t)_{LH}) - (S(t)_{HL} - S(t)_{HH})$$

Each combination of architecture and stopping rule makes a unique prediction for MIC and SIC combination. Modeling results (Figure 3) indicate controls used primarily a *parallel, self-terminating* strategy (3 of 4 participants). ASP participants employed different architectures, but primarily an *exhaustive* stopping rule (3 of 4 participants).

Figure 3. Architecture and Stopping Rule Modeling Results



Control A:
Parallel Self-terminating



Capacity Results
Capacity, a measure of processing efficiency, addresses how a change in work load affects response times. We measure capacity using the following ratio,

$$C(t) = \frac{H_{redundant}(t)}{H_{global}(t) + H_{local}(t)}$$

where H(t) is the integrated hazard function and C(t)=1 indicates unlimited capacity (increase in work load does not change processing rate). For all participants in all groups, C(t) was consistently less than 1, indicating limited processing capacity. Thus, increased work load slowed response times.

None of the participants exhibited extreme limited capacity, as no redundant target reaction time distributions violated the Grice (1984) lower bound on unlimited capacity, parallel processing.

Although all participants exhibited a decrease in response time in redundant target conditions, this was not the result of improved processing efficiency.

Conclusions

Overall, findings from both selective and divided tasks indicate evidence for global advantage in all groups. Results indicate global interference effects across all groups and stimuli types, and local interference effect for all groups on the Arrows tasks. All groups showed faster responses on redundant target trials, suggesting that under divided attention conditions subjects attended to both global and local information. Initial modeling results indicate an important difference in the stopping rule employed by the ASP group compared to the AGEC and ADULT groups. The majority of those with ASP showed exhaustive processing, whereas the majority of the AGEC group demonstrated self-terminating processing. A self-terminating approach is more efficient and thus, should be preferred in this task. Exhaustive processing in ASP has important implications for perceptual processing across domains and may contribute to processing differences previously observed for both non-social and social stimuli.

References

Jolliffe, T., & Baron-Cohen, S. (1997). Are people with autism and Asperger syndrome faster than normal on the embedded figures test? *Journal of Child Psychology, Psychiatry, and Allied Disciplines, 38*(5), 527-534.

Mottron, L., Peretz, I., & Menard, E. (2000). Local and global processing of music in high-functioning persons with autism: Beyond central coherence? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 41*(8), 1057-1065.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology, 9*, 353-383.

Plaisted, K., Swettenham, J., & Rees, L. (1999). Children with autism show local precedence in a divided attention task and global precedence in a selective attention task. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 40*(5), 733-742.

Rinehart, N. J., Bradshaw, J. L., Simon, A. M., Brereton, A. V., & Tonge, B. J. (2000). Atypical interference of local detail on global processing in high-functioning autism and Asperger's disorder. *Journal of Child Psychology, Psychiatry, and Allied Disciplines, 41*(6), 769-778.

Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry, 34*, 1351-1364.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology, 39*(4), 321-359.

Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *The Journal of General Psychology, 127*(1), 67-99.

THE TIME OF OUR LIVES

Mari Riess Jones
Psychology Department, The Ohio State University, 200K Lazenby Hall, 1827 Neil Avenue, Columbus, OH 43210; e-mail: jones.80@osu.edu

In collaboration with Devin McAuley[1], Shayla Holub[1], Heather M. Johnston[2], and Nathaniel S. Miller[1]
[1]*Bowling Green State University,* [2]*The Ohio State University*

Abstract

An entrainment theory of lifespan changes in paced and un-paced tapping was evaluated in participants, ranging in age from 4 to 95 years. Assessments of spontaneous motor tempo and judgments about preferred perceptual tempo converged to show that a preferred time-span (period) of attending slows with age. Variability of tapping also changed systematically with age, consistent with predictions about lifespan changes in entrainment region width and a Restricted Weber Function.

This research has two goals. One goal is to empirically assess performance in a range of sequence timing tasks, including the synchronize-continue task, over the lifespan using a wide range of sequence rates. The second goal was to evaluation adaptations of an entrainment theory of attending to performance in un-paced (spontaneous) and paced (synchronize-continue) tapping tasks.

Generally, entrainment theories assume that the rate (tempo) and rhythm of everyday events engage people on a moment-to-moment basis through attentional synchrony. Entrainment is a natural and ubiquitous process by which an internal periodic activity (i.e., an oscillation) synchronizes with an external event rhythm (Roenneberg, Daan, & Merrow, 2003; Winfree, 2000) Specific models posit a particular relationship between internal (attending) periodicities and event time structure that describe how attending energy may be temporally allocated with reference to a pacing signal of some rate (Jones, 1976b; Large & Jones, 1999; McAuley, 1995).

In this paper we extend the idea of oscillatory attending to address entrainment over the lifespan, by posing these questions: "Do people of different ages rely upon, respectively, different internal periodicities; if so, how do these constrain their ability to track certain events in real time?" One hypothesis, the *preferred period hypothesis,* posits a monotonic progression toward longer internal oscillator periods over the lifespan; thus preferred tempo lengthens with age (Jones, 1976b; Drake, Jones, & Baruch, 2000). A second hypothesis, the *entrainment region hypothesis,* assumes that the range of rates to which an individual can readily entrain also varies with age; the ability to 'lock into' a pacing signal (phase couple) is best in middle age and weak in the very young and very old.

Theoretically, we distinguish a manifest (produced) oscillations, P, from a latent (intrinsic) oscillator, Po. Figure 1A illustrates this distinction. Note that although a pacing signal may force the produced period, P, to adjust (within limits) to event rate, the intrinsic period, Po, remains unchanged. Thus, if a sequence has an overall rate, T, that differs only slightly from Po, the latent oscillator readily promotes an adaptive shift of the manifest period, P, to match T. This results in *period matching* (i.e., P = T). But there are entrainment region limits to period matching (period adaptation) that depend on T i.e., sequence rate, and Po. Within an entrainment region, we predict high accuracy (little difference between average P and T) and low variability (high stability). Figure 2A (solid line) shows a hypothetical accuracy profile for period-matching as a function of sequence; it is a theoretical *detuning function* showing limits of period matching. Outside an entrainment region are increased errors (signed T-P errors) and greater variability (instability) as T becomes remote from Po. This is due in part to the attractive pull of Po (SMT), which will be most evident in continuation (versus synchronized) tapping: P will drift toward Po. In summary, an entrainment view suggests interesting constraints on people's responses to event time (McAuley, 1995).

Two aging hypothesis are sketched in Figure 1B. The Po_1 and Po_2 correspond to intrinsic oscillators with short and long periods, respectively. The preferred period hypothesis implies different latent oscillatory periods for children (Po_1) and adults (Po_2). The entrainment region hypothesis implies that different ranges of accessible periods will obtain for children and elderly adults versus young adults. We assume that Po lengthens with age and that the entrainment region is widest for young/middle-aged adults. These assumptions lead to two predictions. First, *observed detuning functions* (using P –T scores) should shift toward longer rates with age as Po lengthens. Second, the range of accurate tapping rates (low P-T errors) will be greatest for young/middle aged adults. In addition, a *Restricted Weber Function* should describe tapping variability: Weber's Law obtains only within an age-specific entrainment region, as suggested in Figure 2.

Methods

Participants

Eighty-eight children and two-hundred seventeen adults (from 4 to 95 years) were recruited from Northwest and Central Ohio and divided into eight age groups.

Apparatus and tasks

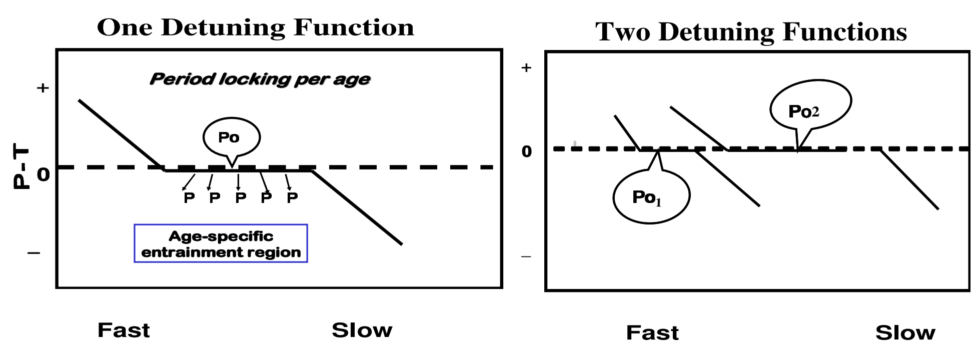


Figure 1: A theoretical detuning function (A) for an entraining oscillator with an intrinsic period, Po as a function of rate, T. The manifest period of the oscillator is P; an entrainment region corresponds to the restricted range of sequence rates where P = T. Two detuning functions (B) based on respectively different intrinsic periodicities; note that the entrainment region is proportionately wider for the oscillator with a longer intrinsic period, Po_2.

Stimuli were generated and responses were recorded by an IBM PC using customized software. Pacing stimuli were isochronous acoustic sequences (50 ms tones of 440 Hz). A response board comprised two copper plates. For the Spontaneous Motor Tapping (SMT)_task participants produced a preferred tempo by tapping for 31 taps at their "most comfortable, natural, rate of tapping". For the Preferred Perceptual Tempo task, participants rated sequences of different tempi that were either faster or slow than their computed SMT. They used a 21 point scale where +10 was "too fast" and – 10 was "too slow" with 0 "just right". For the Synchronize-continue (paced) tapping task, participants first synchronized thirty hand taps with an isochronous rhythm and then continued at that rate for thirty taps. Seven logarithmically spaced rates were (in inter-onset-intervals): T = 150, 225, 337, 506, 759, 1139, and 1709 ms.

Results

Results address: 1. Changes in preferred period, Po; 2. Empirical detuning functions; 3. Tapping variability.

Changes in preferred period

Averages of each participant's four SMT productions were measured using the median of the thirty-interval sequence. Modal SMT distributions shifted from a fast rate (i.e. estimated P_0 of 300 ms) for the 4 to 5 year olds to a fairly slow rate (i.e., a P_0 = 700 ms) for participants over 75 years of age, a trend consistent with the preferred period hypothesis. A linear correlation of age with SMT over all participants was $r = 0.21$, $p < 0.01$. Converging data from perceptual ratings of sequences of different rates show that PPT scores correlated with SMT scores ($r = 0.75$, $p < 0.01$).

Empirical detuning functions in continuation tapping

Accuracy of produced periods, P, at each sequence rate for is plotted in Figure 3 as empirical detuning curves for children (A) and adults (B). Overall, participants were least able to match P with T at the fastest and slowest tapping rates. But age differences were evident. Children showed more negative errors than adults especially at slower rates. Figure 3A shows that error scores for children were positive at the fastest rates (i.e., over-estimates of T) and became increasingly negative at the slowest rates (i.e., under-estimates of T). The youngest children

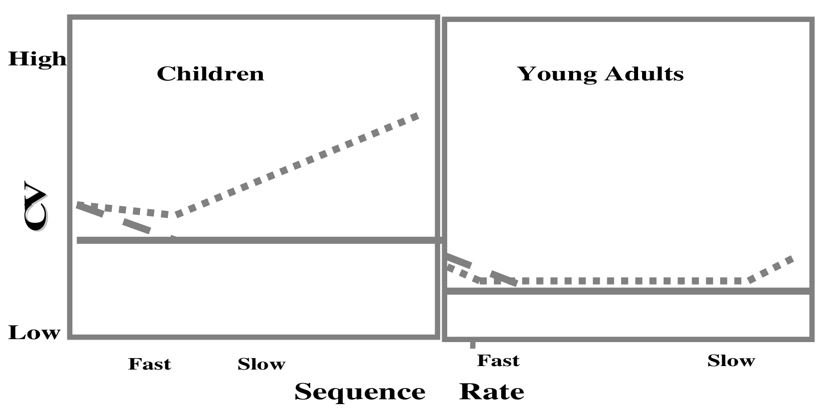


Figure 2: Comparative predictions of interval and entrainment models for coefficient of variability (CV) in continuation tapping of young children (left) and young adults (right). Solid line and dashed lines reflect, respectively, predictions of simple and generalized Weber Laws. The dotted line reflects the Restricted Weber function.

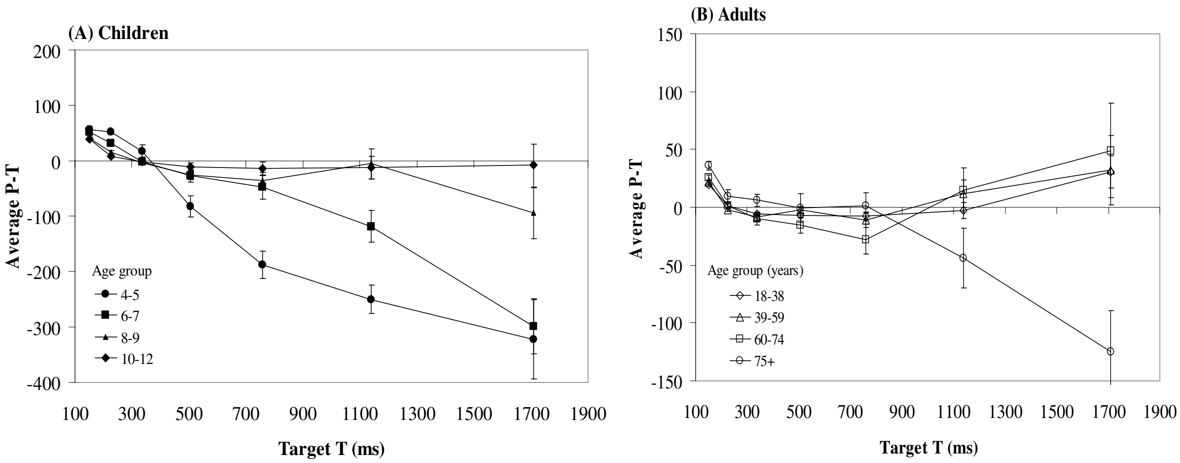


Figure 3: Mean error scores (P – T) in continuation tapping at each sequence rate, T, for children (Panel A) and adults (Panel B).

tended to tap near their preferred tempo (337 ms) at most rates. As predicted, adults were generally accurate and stable over a wide range of rates (relative to children) as shown in Figure 3B. In accord with the entrainment region hypothesis, with increased age, the region of stable tapping tended to shrink; adults ages 60 and older showed significant distortions in accuracy at both the slowest and fastest rates with those over 75 years producing accuracy profiles similar to ones evident in younger children.

Tapping variability and the Restricted Weber Function

Variability data, summarized by CV, speak to Restricted Weber Function. Contrary to the classic version Weber's law, which predicts flat CV functions over rate, the observed CV values were not constant over T for all age groups. Instead, significant departures from Weber's Law (and Generalized Weber's Law) occurred with young children where CV values were disproportionately large at slow event rates as predicted (see Figure 2). These data are consistent with the entrainment region hypothesis.

References

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: attunement, referent period, focal attending. *Cognition, 77*, 251-288.

Jones, M. R. (1976b). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323-355.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119-159.

McAuley, D. J. (1995). *Perception of time phase: Toward an adaptive oscillator model of rhythmic pattern processing*. Indiana University, Bloomington.

Roenneberg, T., Daan, S., & Merrow, M. (2003). The art of entrainment. *Journal of Biological Rhythms, 18*, 110-124.

Winfree, A. T. (2000). *The Geometry of Biological Time* (Second ed. Vol. 12). New York,: Springer-Verlag.

INDUCING LOUDNESS REDUCTION AFFECTS AUDITORY LATERALIZATION

Karen Kelly[1], Yoav Arieh[1] and Lawrence E. Marks[2]
[1]Sensory Information Processing Laboratory, Psychology Department, Montclair State University
[2] John B. Pierce Laboratory and Yale University
ariehy@mail.montclair.edu

Abstract

In Induced Loudness Reduction (ILR), a relatively intense tone reduces the loudness of subsequent weaker tones of the same or nearby frequency. Does inducing loudness reduction also affect auditory lateralization, a task that depends on relative intensity at the two ears, and thus on intensity processing? Listeners were asked to lateralize tones having various differences in dB between the ears, with occasional loud tones (inducers) presented to one ear (treated ear). We found that tones were less likely to be lateralized at the treated ear when they shared frequency with the inducers (Exp. 1) but not when frequencies differed by several critical bands (Exp. 2). We conclude (a) that inducing loudness reduction can affect other intensity-related processes (lateralization) besides loudness; and (b) that ILR arises before or at the neural site where sound location is computed.

Loudness, especially of mid-level tones, is notoriously resistant to simple adaptation. The loudness of an ongoing pure tone declines only a little compared to the decline in subjective intensity observed in other modalities under similar conditions (Scharf, 1983). But loudness is not wholly immune to the history of recent stimulation; in fact, under appropriate conditions, the loudness of a pure tone can decrease by half. Consider the following example: in isolation, a short 2500-Hz target at 60-dB SPL is judged as loud as a 500-Hz comparison at 61 dB. When the same target follows a short 2500-Hz inducer at 80 dB, however, the target is judged to be as loud as a 51-dB comparison, a decline of 10 dB or 50% in sones (Arieh & Marks, 2003a). This impressive reduction in the target's loudness has been labeled loudness recalibration (Marks, 1994) or Induced Loudness Reduction (ILR) (Scharf, Buus, & Nieder, 2002).

Since ILR was first reported (Marks, 1988), a concerted effort has been made to characterize this robust phenomenon (Wagner & Scharf, in press). A few of the central facts about ILR follow (for a more complete review, see Marks & Arieh, in press). (1) ILR transcends a particular measurement procedure. The phenomenon had been observed in loudness matching (Nieder, Buus, Florentine, & Scharf, 2003; Mapes-Riordan & Yost, 1999), magnitude estimation (Marks, 1994; Wagner & Scharf, in press), judgments of loudness differences (Schneider & Parker, 1990), and response time (Arieh & Marks, 2003b). (2) The reduction in loudness represents a decrease in the magnitude of the underlying sensory representation rather than a decisional shift in response criteria (Arieh & Marks, 2003b). (3) ILR is fast to set in but slow to dissipate (Arieh & Marks, 2003a; Arieh, Kelly & Marks, 2005), and (4) ILR is restricted to target tones that fall within the critical band around the frequency of the inducing tone (Marks & Warner, 1991).

Although the cited studies used a variety of experimental procedures, they do have at least one thing in common, they all measured the same dependent variable. That is to say, past research had focused exclusively on the effects of inducers on loudness. It is conceivable, however, that the effects of inducers transcend the perception of loudness per se. Our current understanding of ILR favors a process that happens relatively early in processing auditory intensity and so may modify responsiveness more generally. Thus, it is possible, and in our view even plausible, that inducers attenuate subsequent neural signals of intensity and that this attenuation arises either in the auditory periphery or early in central auditory processing. A straightforward hypothesis follows from this view: inducers will affect most if not all intensity-based auditory processes. This article reports a preliminary test of the hypothesis: Do inducing tones affect other intensity-based auditory phenomena besides loudness?

An attractive candidate for testing the hypothesis is sound localization or, when tones are presented through headphones and therefore 'localized' inside the head, auditory lateralization. The perceived location of a sound relies not only on interaural temporal differences (phase, arrival time) but also on interaural differences in intensity. Given (1) that sounds tend to be heard at the ear where SPL is greater (and where loudness would be greater were the left-ear and right-ear components presented in isolation) and (2) that the effects of inducers (on loudness, at least) transfer moderately, at best, to the contralateral ear (Marks, 1996), we predict that brief, intense inducing tones presented to one ear will depress lateralization at that ear. This prediction was tested in Experiment 1.

Experiment 1

Method

Two men and four women (18-33 years), who reported normal hearing, took part. A Matlab program running on Dell Pentium-IV PC controlled all aspects of stimulus presentation and data collection. Tones were generated by a Tucker-Davis System 3 Real Time processor and were appropriately attenuated (Tucker-Davis PA5 module) and delivered through calibrated TDH-49 headphones mounted in MX41/AR cushions.

Inducing tones lasted 50 msec (5 msec $\cos^2$ gating) and had a frequency of 2500 Hz at a level of 80 dB SPL. Target tones had the same duration and frequency, but we varied the SPLs at the right and left ears (keeping the mean at 60 dB) in order to vary perceived lateral location. For example, one stimulus comprised a 61-dB signal at the right ear presented together with a 59-dB signal at the left. In this case, we expected that listeners would tend to lateralize the tone to the right side. Overall, we created six stimuli, three of which should be lateralized to the right side and three to the left. Table 1 lists the stimuli along with SPLs presented to the right and left ears.

Table 1: Sound levels (dB SPL) presented to the left and right ears

Right ear level	Left ear level	Difference (right – left)	Correct response
62	58	4	Right
61	59	2	Right
60.5	59.5	1	Right
59.5	60.5	-1	Left
59	61	-2	Left
58	62	-4	Left

Each subject participated in three experimental conditions: baseline, right ear induction, and left ear induction. In the baseline condition, subjects were asked to lateralize each tone, chosen from a set of six, to the left or right by pressing the appropriate key on the computer's keypad. The six stimuli were presented a total of 50 times each, and the resulting 300 trials were presented in a different random order for each listener.

The left-ear and the right-ear induction conditions resembled the baseline condition but had one defining difference. At the start of the session and every 15 trials thereafter, a series of 10 inducing tones were played to the left or right ear at a rate of one per second. To minimize carry-over effects, the three conditions were delivered over the course of two sessions, separated by at least by 24 hours. The baseline condition was always presented first, followed by one of the two induction conditions. The remaining induction condition was presented in the second session.

Results

Figure 1 shows the average percentage of 'right ear' responses to each stimulus in each condition. These results show that lateralization was incomplete even when the interaural difference was 4 dB, the maximum tested, an outcome consistent with evidence that complete lateralization requires a difference of approximately 10 dB (Durlach & Colburn, 1978). For example, at baseline, when the SPL was 4 dB greater at either the right or left ear, the tone was lateralized to the right or left about 71% and 74% (100% - 26%) of the time, respectively.

When the inducers were played to the right ear, the average percentage of tones lateralized to that ear decreased substantially compared to baseline. The average of 'right ear' responses across the six stimuli when the inducers were played to the right ear was 30% (versus 45% at baseline). Inspection of Figure 1 suggests that the reduction in 'right-ear' responses is evident in all six stimuli. When the inducers were played to the left ear, the average percentage of responses lateralized to that ear also decreased relative to baseline (or, alternately, the percentage lateralized to the right ear increased). The average of 'left-ear' responses across the six stimuli when the inducers were played to the left ear was 41.7%, (versus 55% at baseline).

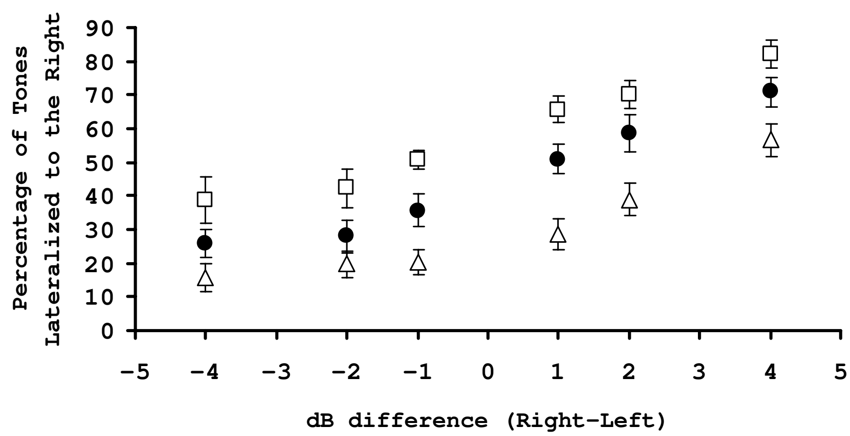


Figure 1: The percentage of right lateralization responses computed for each stimulus condition in the baseline (filled circles), right induction (triangles) and left induction (squares) conditions. Error bars indicate one standard error of the mean.

The results of Experiment 1 strongly suggest that the 80 dB inducers had a substantial effect on lateralization of subsequent 60 dB targets. The tendency to lateralize sounds to the ear just receiving the inducers was reduced, on average by 14.1%. Before attributing the disruption in lateralization to the same mechanism that is underlying ILR, however, we have to rule out a possible interpretation of the results in terms of response bias. It is conceivable that playing a relatively loud sequence of tones to one ear biases the subjects to lateralize subsequent tones to the contralateral ear. Marks and Warner (1991) reported that inducers affect the loudness only of tones that lie within the inducers' critical band. Consequently, the hypothesis that inducers reduce neural intensity signals within a critical band led us to predict that 500-Hz inducers should have little or no effect on lateralization of 2500-Hz targets. This prediction was tested in Experiment 2.

Experiment 2

Method

Three men and three women (ages 18-42 years), who reported normal hearing, participated in Experiment 2. In this experiment, the frequency of the inducer tones was 500 Hz instead of 2500 Hz. Otherwise, the procedure and stimuli were identical to those of Experiment 1.

Results

Figure 2 shows the average percentage of 'right ear' responses given to each stimulus in each condition. In contrast to the differences among the lateralization functions in Experiment 1, the three functions in Experiment 2 nearly overlap. This parity across conditions suggests that the 500-Hz inducers in Experiment 2 did not affect lateralization significantly. The grand means of lateralization responses to the right ear were 51.4%, 51.9% and 47.1% for the baseline, right-ear induction, and left-ear induction conditions, respectively. Thus, in Experiment 2, hearing a loud tone at one ear did not induce the subjects to lateralize subsequent weaker tones to the other ear. These results comfortably rule out a simple response-bias interpretation of the effect of the inducers on lateralization.

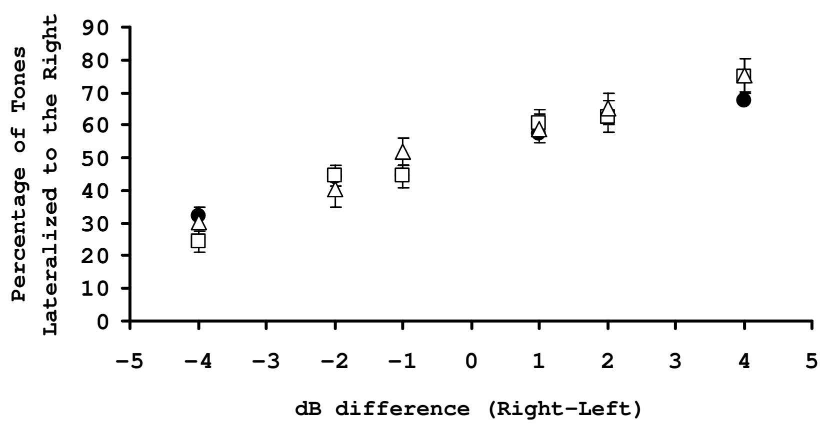


Figure 2: The percentage of right lateralization responses computed for each stimulus in the baseline (filled circles), right induction (triangles), and left induction (squares) conditions. Error bars indicate one standard error of the mean.

Discussion

In this study, we used the stimulus conditions that produce ILR to show that presenting a sequence of intense inducers to one ear decreases the tendency to lateralize weaker targets to that ear. We conclude that the aftereffects on auditory perception of presenting short, intense tones are more widespread than heretofore understood. Brief, intense inducers not only affect the perception of loudness of subsequently presented weaker tones but also disrupt sound lateralization. Because the shifts in lateralization produced in these experiments depended solely on relative sound intensity at the ears, we assume that the inducers reduced the magnitude of the 'neural intensity signal' in the treated ear. Thereby modifying the difference between the neural signals derived from the two ears at the site where lateral position is computed, 'favoring' the untreated ear. Implicit in this view is the supposition that the effect of inducers on loudness, or ILR, is just one example of a much wider set of inducer-generated auditory aftereffects. The current results lend a support to the notion that brief, intense tones cause a general decrease in responsiveness to subsequent auditory signals and thereby modify the neural signals that are pertinent to, say, decisions regarding lateralization.

As far as we know, this is the first report to document the effect of brief inducing tones on auditory processes other than loudness perception. The results show that presenting brief, intense tones may have several functional consequences for intensity-dependent auditory behavior, not just reduction in loudness, and suggest that these consequences may appear in still other auditory processes, such as stimulus identification and distance estimation. Finally, the results imply that the very terms that have been used to label the effect on loudness – induced loudness reduction, loudness recalibration – refer to only one facet of the phenomenon. Tentatively we offer the more general term Post Transient Suppression.

Acknowledgement

Preparation of this article was supported in part by grant DC00271-18 from the National Institutes of Health to the third author.

References

Arieh, Y., & Marks, L.E. (2003a). Time course of loudness recalibration: Implications for loudness enhancement. *Journal of the Acoustical Society of America*, *114*, 1550-1556.

Arieh, Y., & Marks, L.E. (2003b). Recalibrating the auditory system: A speed-accuracy analysis of intensity perception. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 523-536.

Arieh, Y., Kelly, K., & Marks, L. E. (2005). Tracking the time to recovery after induced loudness reduction. *Journal of Acoustical Society of America*, 117, 3382-3384.

Durlach, N. I ., & Colburn, H. S. (1978). Binaural phenomena. In E. C. Carterette & .M. P. Friedman (Eds.), *Handbook in Perception* (Vol. 4, pp. 365-466), Hearing. New York: Academic Press.

Mapes-Riordan, D., & Yost, W. A. (1999). Loudness recalibration as a function of level. *Journal of Acoustical Society of America*, *106*, 3506–3511.

Marks, L. E. (1988). Magnitude estimation and sensory matching. *Perception & Psychophysics*, *43*, 511–525.

Marks, L. E. (1994). "Recalibrating" the auditory system: The perception of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 382–396.

Marks, L. E. (1996). Recalibrating the perception of loudness: Interaural transfer. *Journal of the Acoustical Society of America*, *100*, 473-480.

Marks, L. E., & Arieh, Y. (in press). Differential effects of stimulus context in sensory processing. *European Review of Applied Psychology*.

Marks, L. E., & Warner, E. (1991). Slippery context effect and critical bands. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 986–996.

Nieder, B., Buus, S., Florentine, M., & Scharf, B. (2003). Interactions between test and inducer tone durations in induced loudness reduction. *Journal of the Acoustical Society of America, 114,* 2846-2855.

Scharf, B. (1983). Loudness adaptation. In J. V. Tobias & E. D. Schubert (Eds.), *Hearing: Research and theory* (Vol. 2, pp. 1–56). New York: Academic Press.

Scharf, B., Buus, S., & Nieder, B. (2002). Loudness enhancement: induced loudness reduction in disguise? (L), *Journal of the Acoustical Society of America*, *112*, 807-810.

Schneider, B., & Parker, S. (1990). Does stimulus context affect loudness or only loudness judgment? *Perception & Psychophysics*, *48*, 409–418.

Wagner, E., & Scharf, B. (in Press). Induced loudness reduction as a function of exposure time and signal frequency. *Journal of the Acoustical Society of America*.

RAVING ABOUT RAVENS: MODELLING SPEED-ACCURACY IN INTELLIGENCE TESTS

Diana Eugenie Kornbrot

Psychology Department, University of Hertfordshire d.e.kornbrot@herts.ac.uk.

Abstract

The effect of time pressure on performance on intelligence tests is a long standing problem. In this study a computerised version of the Ravens Advanced Progressive Matrices was administered using 3 different forms of instructions: control, speed pressure, and accuracy pressure. Analyses used Rasch measures of participant ability and item difficulty, and the time each participant took to solve each problem. Raw scores were, surprisingly, more useful than Rasch measures. The time pressure group were faster but scored less well than the other two groups. Raw score had a small but significant correlation with total test time. Brighter participants took less time for easy items, but more time for hard items, which were both slower and more variable than easier items. Mean and SD were more consistent for total time than for either correct or error time. Effective models will need to incorporate these diverse results.

The Raven's Progressive Matrices is one of the most successful culture fair tests of cognitive functioning (Raven, 1956; Raven, Court, & Raven, 1988; Salthouse, 2005). Like all tests and examinations, performance is potentially subject to time pressures. The first, and most practical aim, of this study is to make use of computerised administration to investigate those time pressures at the item level. A more theoretical aim is to produce a model of performance for this complex task that parallels models of simple perceptual decision making. The aim is to generate for each individual one parameters that corresponds to rate of accumulation of useful information, and one parameter that corresponds to speed bias (like the separation of barriers in a random walk). There have been many studies comparing performance on elementary tasks such as simple and choice reaction time with performance on some version of the Raven's. Typically, such studies use mean reaction time and d' or Luce's choice, ln(η) for the elementary processes, but only raw score for the Raven's (Beh, Roberts, & Prichard Levy, 1994; Fink & Neubauer, 2001; Salthouse, 2005). Even total time, easily obtained with a stopwatch, is rarely used as a peformance measure for the Raven's. Studies looking at time for individual items, requiring computerised administration are even rarer. There are however, several studies that analyse Ravens results using the Rasch model (Alderton & Larson, 1990; Forbes, 1964; Gallini, 1983; Green & Kluever, 1992; Pitariu, 1986). The two or three parameter Rasch models give participant ability and item difficulty measures that depend on logits of probabilities and are hence very similar to the bias and sensitivity measure of choice model.

The present study measures time per item for each person for each item. It explores four measures of individual person performance: raw score, and one, two and three parameter ability. The relation between time per item for the 36 items and individual performance is explored using exploratory principal components analysis. The relation between item difficulty (proportion of people giving correct response or one, two or three parameter item difficulty) to time per item was also explored.

Method

Participants & Design

60 female and male participants, aged from 18 to 72 were recruited from the university population. Each was randomly allocated to one of 3 groups: control, time pressure or accuracy pressure.

Apparatus and Materials

The apparatus was a computerised version of the Ravens© Advanced Progressive Matrices, comprising 12 practice and 36 test items, presented on a MAC computer. The instructions preceding the item administration included the following text presented on the computer screen.

> This is a test of observation and clear thinking. On each of the screens that follow you will see a pattern with a piece cut out of it. Look at the pattern. Think what piece is needed to complete it correctly both along and down must look like. Then find the right piece out of the eight bits shown below. When you think you have found the right piece, click on the piece with the mouse. Your selected answer will be displayed alongside the problem number on the right-hand side of the screen. If you make a mistake or want to change your answer, click on the appropriate piece and your answer will be updated. Select each problem from the list displayed on the right-hand side of the screen by clicking on the appropriate number. You can complete the problems in any order you like. However, the problems are simple at the beginning and get harder as you go along. There is no catch. If you pay attention to the way the answers to the easy problems are found you will find the later ones less difficult.
>
> PRACTICE PHASE
> There are 12 practice problems. Try each in turn. The first problem will be demonstrated for you. Then complete the 11 remaining practice problems. When you have completed all 12 problems click the 'Finished' button and await further instructions from the experimenter.
>
> TEST PHASE
> *Group specific instructions* preceding the following text to all participants
>
> You will be presented with 36 problems. Do not miss any out. If you are unsure of your answer, guess as guesses are sometimes right. If you get stuck, move onto the next problem and then come back to the one you had difficulty with. When you have completed all 36 problems click the 'Finished' button and inform the experimenter.
>
> GROUP SPECIFIC INSTRUCTIONS
> Control Group,1: *None*
>
> Time Pressure Group 2: You are required to perform a general intelligence test. We are interested in how successfully you can complete the test. Success will not only be measured by the number of correct answers obtained, you will also receive credit for the speed with which you complete the task. The participant with the highest overall score based on the number of correct responses <u>and</u> the time to complete the task will receive a prize.
>
> Accuracy Pressure Group, 3: You are required to perform a general intelligence test. We are interested in how successfully you can complete the test. Success will be measured by the number of correct answers obtained. The participant with the highest overall score based on the number of correct responses will receive a prize

Procedure

Each participant signed the consent form and was then seated in front of the computer and informed of the task, as describe in the reference manual. This was followed by computerised presentation of instructions followed the practice and test administration. A conspicuous clock was present on the screen throughout the time pressure condition. There was a pause for participants to relax and ask questions between the practice and test phases of the experiment.

Results

One, two and three parameter Rasch analyses were performed on the frequency of a correct response by each participant to each item. Each analysis gave an item difficulty score or each item and an ability score for each participant. The performance of these measures can then be compared with raw score for each participant and proportion correct for each item. All statistical tests were conducted at the 95% confidence level, with confidence levels in parentheses, as appropriate.

Group and Participant Performance

Table 1 shows means and standard deviations for test score for all 3 groups. ANOVA gave a significant effect of condition for both score, $F(57,2) = 8.1$, $p = .001$, effect size partial $\eta^2 = .221$ and total time, $F(57,2) = 3.3$, $p = .043$, partial $\eta^2 = .10$. Since there was a predicted order for both score and time such that time pressure < control < accuracy pressure, post hoc tests were conducted without multiple comparison correction. With this proviso, the time pressure group was significantly less accurate than the other two groups, and also took significantly less time,. There were no significant differences between the control and accuracy pressure groups.

Table 1
Mean & standard deviation for raw Raven's score and total time to complete the test for all group

	control N=20		time pressure N=21		accuracy pressure N=19	
	mean	sd	mean	sd	mean	sd
test score/36	23.5	4.3	20.0	5.8	26.2	4.4
total time, mins	52.6	16.0	42.1	13.2	54.6	20.0

Alternative performance measures gave similar effect sizes to those obtained for the raw score. Partial eta squared = .224 for logit(probability correct), .209 for probability correct times logit(probability correct); and .198 for information favouring correct response. Group differences did not show up at all using ability measures from any of the Rasch models.

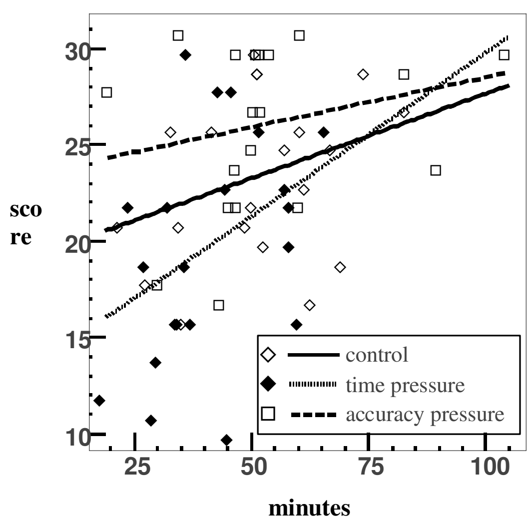


Figure 1. Raven's score as a function time to complete test for 3 groups.

Figure 1 shows score as a function of time to complete the test for all 3 groups. An ANCOVA with score as response variable and time, condition and time*conditions as explanatory variables gave a main effects of condition, $F(2,54) = 8.6$, $p = .001$, partial $\eta^2 = .24$; and time, $F(1,54) = 5.6$, $p = .022$, partial $\eta^2 = .09$; but no significant interaction term. The model with only main effects gave a slope of .090 confidence limits (.014, .166) and intercepts for control = 18.8 confidence limits(16.7, 20.9); time pressure = 16.2 with confidence limits(14.1, 18.3); accuracy pressure = 21.3 confidence limits(19.3, 23.4). So participants scored .9 points for each extra 10 minutes spent on the problem. Regression of the Rasch ability measures on total time in minutes were also conducted. Results were similar to those for raw scores but accounted for slightly *less* variance, so they are not shown

The relation between measures of correctness and time spent on each separate question was investigated by performing exploratory principal components analysis with time spent on each of the 36 problems and some measure of ability. Four separate analyses were conducted using the 4 possible measures of ability, total score, ability1, ability2 and ability3, where the digit following ability indicates the number of parameters in the generating Rasch model. The results were essentially identical for all ability measures. Table 2 shows the varimax rotated loadings for the solution using total score as the ability measure. Two components accounted for 41% of the variance. The first loaded *positive* on score and more than .45 on time spent on the more *difficult* items 19 onwards. The second loaded *negatively* on score and more than .45 on time spent on the *easier* items 1-17. Exceptions were problems 8 & 18, loading nearly equally on both components. It appears that abler participants spend less time on easy items but more time on difficult items.

Table 2. *Rotated principal components loadings for score and time to complete each problem*

Variable	score	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
Component 1	.64			.34		.36			.48										.50
Component 2	-.48	.51	.54	.62	.64	.60	.72	.69	.43	.62	.70	.66	.38	.40	.60	.53	.48	.62	.49
Variable		T19	T20	T21	T22	T23	T24	T25	T26	T27	T28	T29	T30	T31	T32	T33	T34	T35	T36
Component 1		.50	.63	.33	.43	.71	.40	.67	.68	.70	.73	.61	.78	.66	.79	.57	.54	.67	.65
Component 2					.36														

Item Performance

The item difficulty level could be measured in 5 ways: by problem number (the original validation has problems in order of difficulty), by proportion of participants getting correct response, Pcor, and by the 3 different Rasch difficulties. Figure 2 shows mean of all times and sd off all times as function of number correct, together with sd as a function of mean. Each graph has 36 points generated by the 36 different items. Adjusted r^2 were as follows: regression of mean on number correct, $r^2 = .820$; for regression of SD on number correct, $r^2 = .822$; for SD on mean, $r^2 = .937$.

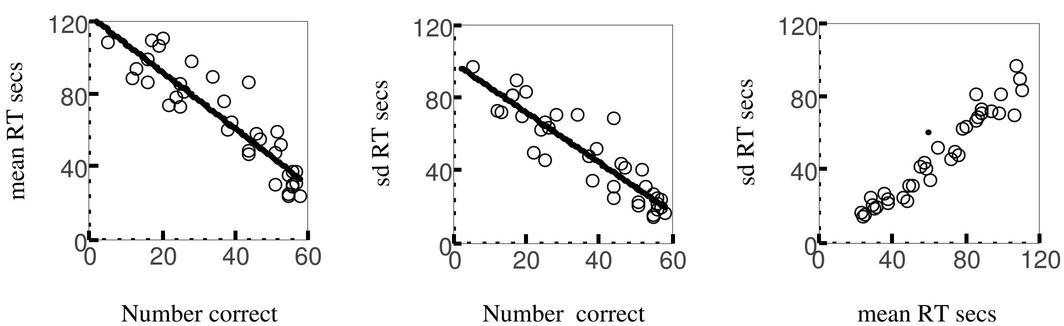


Figure 2. Mean RT as function number correct (left panel); and SD RT as function number correct (middle panel); SD RT as a function of mean RT (right panel).

The mean and s.d of total time, time for correct response and time for error response for each item was regressed separately on each of the five measures of item difficulty. Higher linear correlation coefficients were found for total time than for either correct or error response time. Furthermore there was no significantly difference between mean error and mean correct times. The highest correlations were obtained for regression on number correct for both mean and standard deviation. Means and standard deviations were highly correlated.

Summary

The general level of performance is much a would be anticipated for a university community (Bors & Stokes, 1998; Bors & Vigneau, 2001; Paul, 1985). There was a quite substantial effect of time pressure on both performance and total time taken. However, there was no difference between the standard administration control group and the accuracy pressure group. This is encouraging suggesting that left to themselves people do indeed opt of the accuracy strategy. Nevertheless, in the time pressure group only 1 person took more than an hour whereas in the other two groups combined 11/39, i.e. more than 25% took more than 1 hour. Standard administration is often limited to 1 hour. Clearly this can cause underestimation of performance.

There was a small positive correlation between time and performance, the same for all groups. Participants' performance was better by nearly 1 whole point for each extra 10 minutes spent. However, the direction of causality is unknown. Brighter participants may choose to take longer.

The results of the principal components analysis are new and interesting. Factor analysis of Raven's item themselves shows only a single factor. Nevertheless, abler participants are faster on easy items, but slower on difficult items. This may explain some of the relation between total time and raw score. The less able participants may guess on the difficult items, which is less time consuming than problem solving.

The relation between item difficulty, as measured by participant success rate, to time per item is also interesting. People spend longer on the more difficult items, but performance is also more variable as shown in Figure 2.

In this study raw scores outperformed Rasch measures of ability on all fronts. Raw person ability scores showed larger group difference effect sizes, stronger correlations with total reaction time and more variance accounted for in principal component analysis. Raw item difficulty measures showed stronger correlations with time per item than Rasch difficulty measures. This is good news for simplicity of analysis. There is a very simple message. Stick with raw scores.

However, the poor performance of the Rasch measures is disappointing, and in my view surprising, from the perspective of modelling performance. I had anticipated that Rasch measures, similar to parameters of choice model would be more closely related to time measures. From this perspective it is interesting that total time is a more reliable measure than either correct or error time, as would be predicted by relative judgement theory. Unfortunately this might be an artefact of the limitation that some error or correct times were based on small numbers of observations, while total time was always based on all 36 items. In addition, information measure of performance were just as bad as the (related) Rasch measures in terms of variance accounted for in the various analyses.

In summary, Raven's matrices remain an excellent tool for assessing cognitive ability. Raw scores seem better than any transformation. Time pressure, i.e. a fixed time limit is not a good idea. Modelling of processes involved in solving the Raven's has a very long way to go. Such modelling will need to take into account the new finding that more able participants are only faster on easy items. For harder items, they appear to strategically increase the time spent to accommodate the fact that more information is needed to solve the problems.

Acknowledgements

Thanks to Campbell Thomson & McLaughlin Ltd, who were extremely helpful and permitted the production of the computerised version of the test. Raven's Advanced Progressive Matrices Test. Raven Progressive Matrices are stored digitally for this study with the permission of J.C. Raven Ltd. Advanced Progressive Matrices © John C. Raven 1943, 1947, 1962. © J.C. Raven Ltd 1976. All rights in Raven Progressive Matrices are reserved, and they may not be reproduced copied down-loaded stored in a data base or published in whole or in part in any form except with the express permission of the copyright owners. Enquiries about copyright in the RPM may be made to Campbell Thomson & McLaughlin Ltd, 1 King's Mews, London WC1N 2JA, tel: 020-7242-0958, fax: 020-7242-2408. Thanks also to Jed Everitt who programmed the computerised version of the test and to David Wellsted who tested the main bulk of participants.

References

Alderton, D. L., & Larson, G. E. (1990). Dimensionality of Raven's Advanced Progressive Matrices items. Educational and Psychological Measurement, 50, 887–900.

Beh, H. C., Roberts, R. D., & Prichard Levy, A. (1994). The relationship between intelligence and choice reaction time within the framework of an extended model of Hick's Law: A preliminary report. Personality and Individual Differences, 16(6), 891-897 Record 812 of 813 in PsycINFO 1992-1994.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. Educational and Psychological Measurement, 58(3), 382-398 Record 388 of 313 in PsycINFO 1995-1997.

Bors, D. A., & Vigneau, F. (2001). The effect of practice on Raven's Advanced Progressive Matrices. Learning and Individual Differences, 13(4), 291-312.

Fink, A., & Neubauer, A. C. (2001). Speed of information processing, psychometric intelligence: And time estimation as an index of cognitive load. Personality and Individual Differences, 30(6), 1009-1021 Record 1006 of 1013 in PsycINFO 1998-1999.

Forbes, A. R. (1964). An item analysis of the Advanced Matrices. British Journal of Educational Psychology, 34, 223-236.

Gallini, J. K. (1983). A Rasch analysis of Raven item data. Journal of Experimental Education, 52(1), 27-32.

Green, K. E., & Kluever, R. C. (1992). Components of item difficulty of Raven's matrices. Journal of General Psychology, 119(2), 189-199.

Paul, S. M. (1985). The advanced Raven's Progressive Matrices: normative data for an American university population and an examination of the relationship with Spearman's g. Journal of Experimental Education, 54, 95–100.

Pitariu, H. (1986). Item analysis and standardization of advanced progressive matrices (APM) / Analiza de itemi si standardizarea matricilor progresive avansate (MPA). Revista de Psihologie, 32(1), 33-43.

Raven, J. C. (1956). Progressive Matrices, sets A, B, C, D and E. London.

Raven, J. C., Court, J. H., & Raven, J. (1988). Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: Advanced Progressive Matrices, sets I and II. (1988 ed ed.). Oxford: Oxford Psychologists Press.

Salthouse, T. A. (2005). Relations Between Cognitive Abilities and Measures of Executive Functioning. Neuropsychology, 19(4), 532-545 Record 532 of 513 in PsycINFO 2004 Part B.

RESPONSE TIME DISTRIBUTIONAL ANALYSIS OF SYMBOLIC COMPARISON PHENOMENA

Craig Leth-Steensen (craig_leth_steensen@carleton.ca) Carleton University

Abstract

For this study, a large set of symbolic comparison data for an artificially induced, six-term linear ordering was obtained from 3 individuals who each performed comparisons over the course of 12 experimental sessions. Having such a large number of observations per pair, per form of the comparative instruction, per participant allowed, for the first time within a symbolic comparison paradigm of this nature, for an examination of the effect of extended amounts of practice on the pattern of observed RTs for individual participants as well as for an examination of the nature of the within-subject RT variability in terms of the shapes of the RT distributions obtained in each experimental condition from each individual participant.

Consider a set of symbolic items ordered in memory according to some unidimensional attribute continuum that is either physically- (e.g, height) or psychologically-based (e.g., likeability). The paired comparison of such items has been the subject of much study over the last 35 years or so (Leth-Steensen & Marley, 2000). As of yet, though, no consensus viewpoint regarding the nature of the representations and mental mechanisms underlying the comparative processing of such order information has been forthcoming.

Group Versus Individual Data

Much research in sensory, perceptual, and especially cognitive psychology has involved the collection of a relatively small amount of data from each of a large number of individuals and then an examination of the averaged data. Although, such averaging is generally regarded as a useful way to remove variation due to individual differences, there are several reasons why the collection of a larger set of observations from several individual participants can also be deemed desirable. Two of them involve concerns as to the effects of averaging data (c.f., Ashby, Maddox, & Lee, 1994) and the effects of qualitative performance changes with practice (c.f., Logan, 1988; Rabbitt & Banjerie, 1989), respectively. The other arises from the view that it has become necessary to obtain empirical measures that can describe performance at a more fine-grained level than is provided by mean response time (RT) and percent correct (PC; Ratcliff & Murdock, 1976). One way to obtain such measures is from quantitative descriptions of the shapes of the RT distributions. These measures can often provide important new information concerning the nature of the processing involved in the performance of the task in question. Such information can be used to evaluate the usefulness of any existing theories which claim to be able to account for that performance by providing a fuller set of empirical constraints on the predictions that any relevant theory must then provide (e.g., Mewhort, Braun, & Heathcote, 1992). Although RT distributional grouping techniques are available (e.g., Vincent averaging; Ratcliff, 1979), the best way to obtain such measures is by collecting a large amount of data from a few individuals.

The Ex-Gaussian Distribution

The ex-Gaussian distribution can provide useful quantitative measures of the distributional properties of a set of RTs (Heathcote, 1996; Luce, 1986). The ex-Gaussian model assumes that each RT can be represented as the sum of a normally distributed and an independent exponentially distributed random variables, and therefore, that the full distribution of RTs can be characterized as a convolution of the normal and exponential distribution functions. Parametrically, the ex-Gaussian distribution has three constituents: μ and σ, which respectively describe the mean and standard deviation of the normal component (which itself could represent the sum of a number of normal components); and τ, which describes the mean of the exponential component (which must represent the contribution of a single dominant exponentially distributed component). The initial theoretical focus on the ex-Gaussian distribution (e.g., Hohle, 1965) involved attempts to identify each of these two mathematical components with the time courses of various stages of processing such as the encoding and motor stages, or the decision stage (with not a lot of success; Luce, 1986; see also Heathcote, Popiel, & Mewhort, 1991). However, in a more practical sense, whenever empirical RT distributions are fit with the ex-Gaussian, the μ and τ parameters provide excellent quantitative measures that reflect both the location of the leading edges (or modes) of the distributions and the size of their tails (which provides a rough indication of the degree of positive skew), respectively.

Method

Subjects

Three right-handed, University of Illinois graduate students (two males and one female) performed 12 separate 1-hour sessions for which they were each paid $116 (US) in total.

Stimuli and Apparatus

The comparison stimuli were six 3-letter male names ("Bob", "Don", "Jim", "Mel", "Pat", and "Ted"). The participants were told only that these names corresponded to six imaginary "individuals" who differed in height. During each comparison trial, two of the name stimuli were presented side by side (horizontally) in the centre of a computer screen.

Procedure

Each session of the experiment involved two phases. In Phase 1 (the training phase), each participant was presented with each of the five comparison pairs consisting of the stimuli that were adjacent to each other in the relevant ordering (i.e., the Split 1 pairs). Each comparison trial began with a blank computer screen for 1000 ms, then the relevant comparative instruction for that trial ("Taller?" or "Shorter?") was displayed for 1000 ms. When the names appeared, the participants chose the name that they thought corresponded to either the shorter or the taller of the two individuals in the comparison pair. Both the comparative instruction and the pair of names remained on the screen throughout the trial. The participants responded by pressing either the left or the right button on a response box with the index fingers of their left or right hand. After each response, accuracy feedback was immediately provided. The participants could examine this feedback for as long as they liked and initiated the next learning trial by pressing a third button. Each experimental block in Phase 1 consisted of 20 comparison trials: 5 (pairs) x 2 (instruction type) x 2 (left-right stimulus presentation order). The participants continued the training phase until they responded correctly to all of the comparison pairs within a full block of 20 trials. The participants were retrained at the beginning of each of the following experimental sessions.

In Phase 2 (the test phase) the procedure was very similar. However, all of the 15 possible pairs of names were now included in the stimulus set and no feedback was provided. Hence, each experimental block in Phase 2 consisted of 60 randomly presented comparison trials: 15 (pairs) x 2 (instruction type) x 2 (left-right stimulus presentation order). In the first session, each participant performed 6 blocks of Phase 2 trials. In the second and third sessions, each participant performed 8 and 10 blocks, respectively, of Phase 2 trials. From the fourth to the eighth sessions, each participant performed 12 blocks of Phase 2 trials in each session. From the ninth to the twelfth sessions, each participant performed 14 blocks of Phase 2 trials in each session. During Phase 1, the participants were instructed to emphasize accurate responding. However, during Phase 2 the participants were instructed to emphasize speed (without sacrificing accuracy).

Results

In general, none of the three participants found it difficult either to learn the training pairs in the first session or to remember them at the beginning of later sessions. As expected, overall RT also decreased with practice for all three participants and began to level out (i.e., asymptote) by the third session. Figures 1, 2, and 3 show the mean RT data for correct responses (collapsed over the left-right order of stimulus presentation) corresponding to the responses made by each participant to each of the comparison pairs with each of the comparative instructions in the test phases of Sessions 3-12. All of the corresponding PC data are also given in each of these figures. Error bars corresponding to two standard errors are provided with each data point. The number of observations collected per pair, per instruction, per participant was 252 in Sessions 3-12. These figures also provide the results of the ex-Gaussian analyses of the full distributions of correct RTs associated with each of the mean RT data points. Fits of the ex-Gaussian model were obtained using the quantile maximum likelihood (QML) procedure of Heathcote, Brown, and Mewhort (2002). The error bars in those figures correspond to two standard errors (provided by QML) away from the fitted parameter estimates.

Discussion

Figures 1-3 clearly indicates that, for the most part, all three of the major symbolic comparison RT effects (i.e., distance, end, and semantic congruity) can be found in the mean correct data of all three of these well-practiced subjects. For example, the RT data for Participant 1 represents an almost "picture-perfect" display of the kind of RT data that would be obtained in any set of averaged group results for these types of comparisons. Namely, bowed end effects at all splits, distance effects that occur mainly due to graded decreases in comparison RTs to the middle pairs, and full semantic congruity cross-over effects that are somewhat funnel-shaped due to the presence of a lexical markedness effect (i.e., faster overall response times to the "Taller" instruction).

With respect to the corresponding ex-Gaussian results, several interesting aspects of responding are revealed by these measures. First, it seems quite clear (especially for Participants 1 and 3) that the semantic congruity effect is manifested mainly in the leading edge of the RT distributions as reflected in the results for μ. Second, both the RT differences between the end and middle pairs and the decreases in mean RT with split for the middle pairs (especially for Participants 1 and 2) also seem mainly due to decreases in the leading edge of the RT distributions as reflected in the results for μ. Third, qualitative differences in the overall variability of the RT distributions between the end and middle pairs (again, especially for Participants 1 and 2) are reflected in highly inflated values of σ for most of the middle pairs. Fourth, for the one participant (Participant 1) who displayed clear lexical markedness effects, these effects are present only in the tails of the RT distributions as reflected in the results for τ. Finally, there is a tendency (which is especially evident in the Split 1 data) for the pairs that were responded with quite low accuracy to be associated with higher values of τ (obtained from ex-Gaussian fits to correct responses).

References

Ashby, F.G., Maddox, W.T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity choice model. *Psychological Science*, *5*, 144-151.

Heathcote, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, and Computers, 28,* 427-445.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin and Review, 9,* 394-401.

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin, 109,* 340-347.

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology, 69,* 382-386.

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, *107*, 62-100.

Logan, G. D. (1988). Toward an instance theory of automization. *Psychological Review, 95,* 492-527.

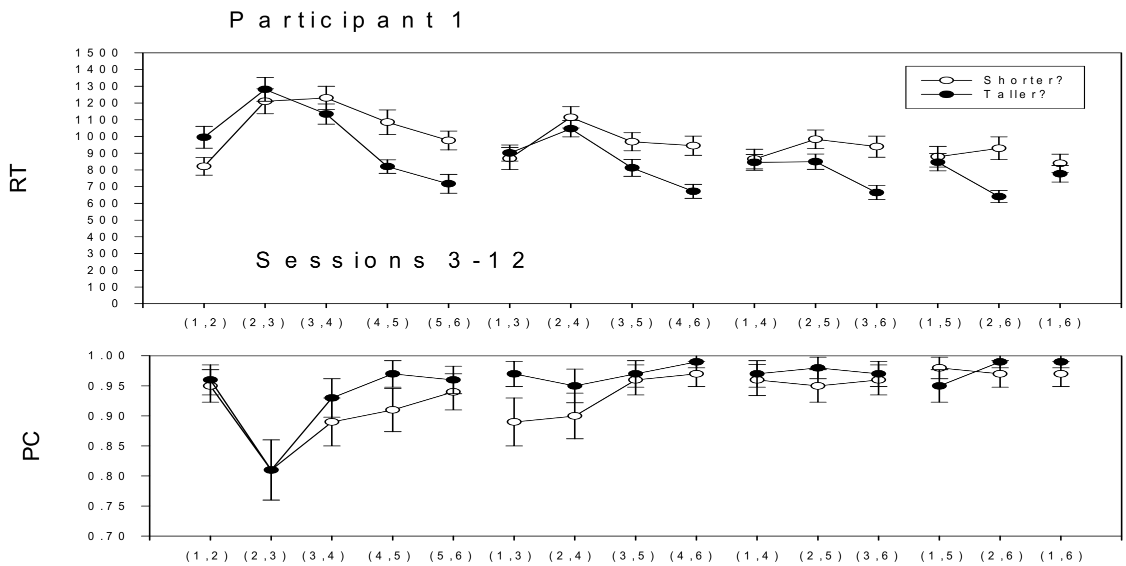
Luce, R. D. (1986). *Response times their role in inferring elementary mental organization.* New York: Oxford.

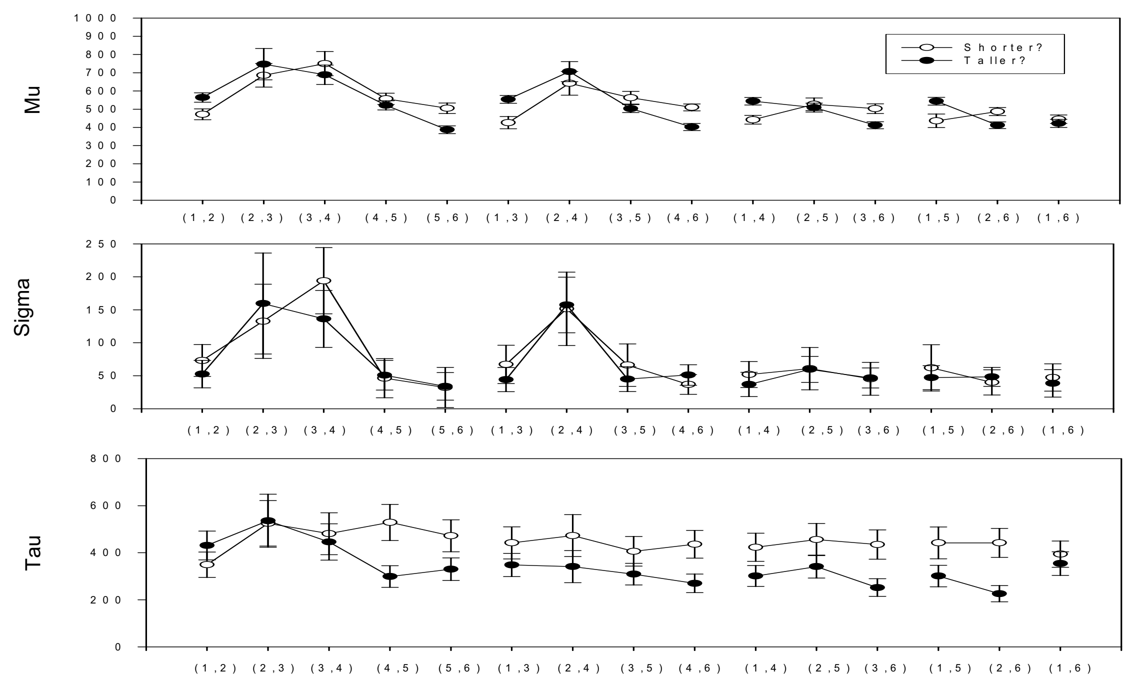
Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response time distributions and the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 872-882.

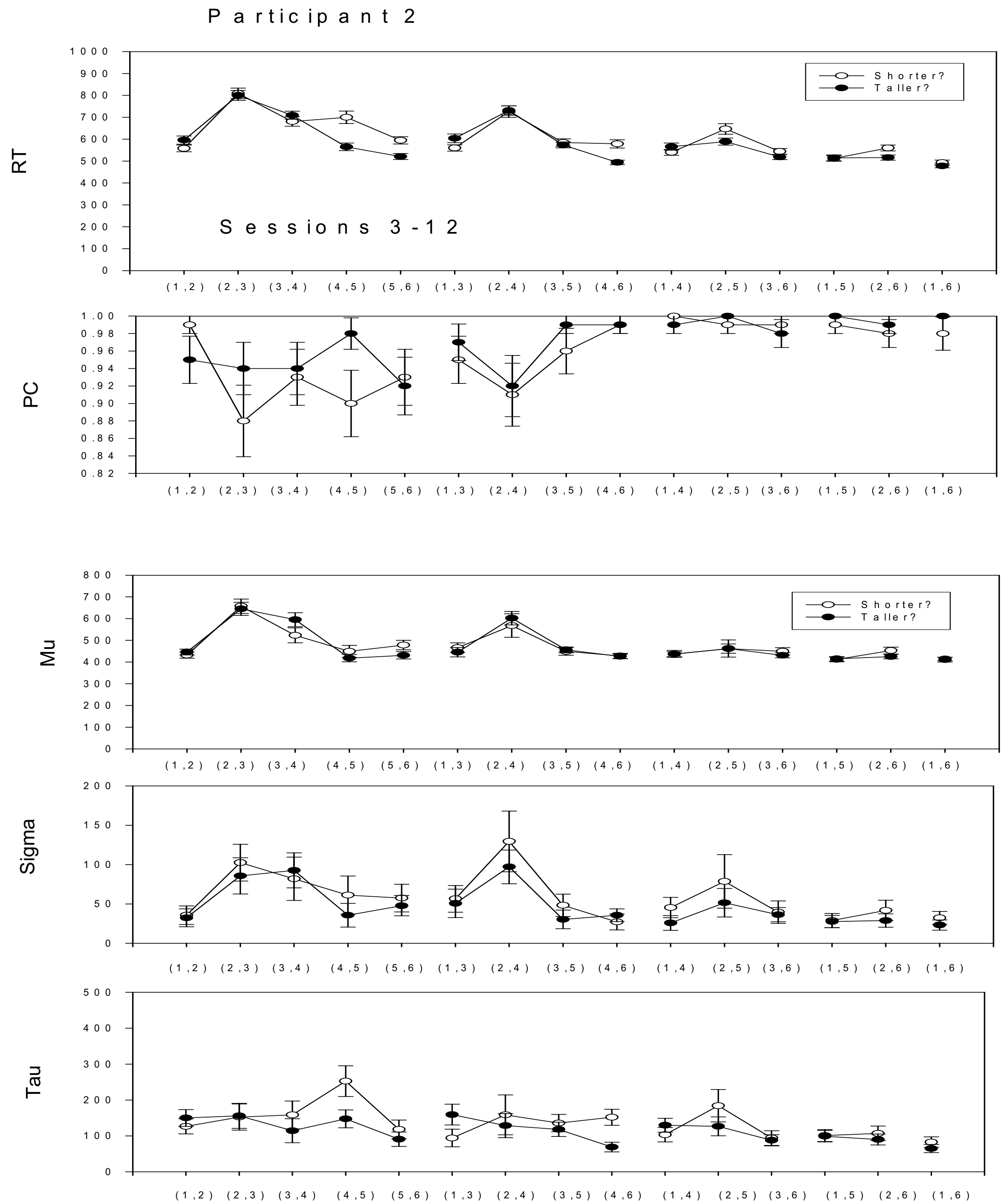
Rabbitt, P., & Banerji, N. (1989). How does very prolonged practice improve decision speed? *Journal of Experimental Psychology: General, 118,* 338-345.

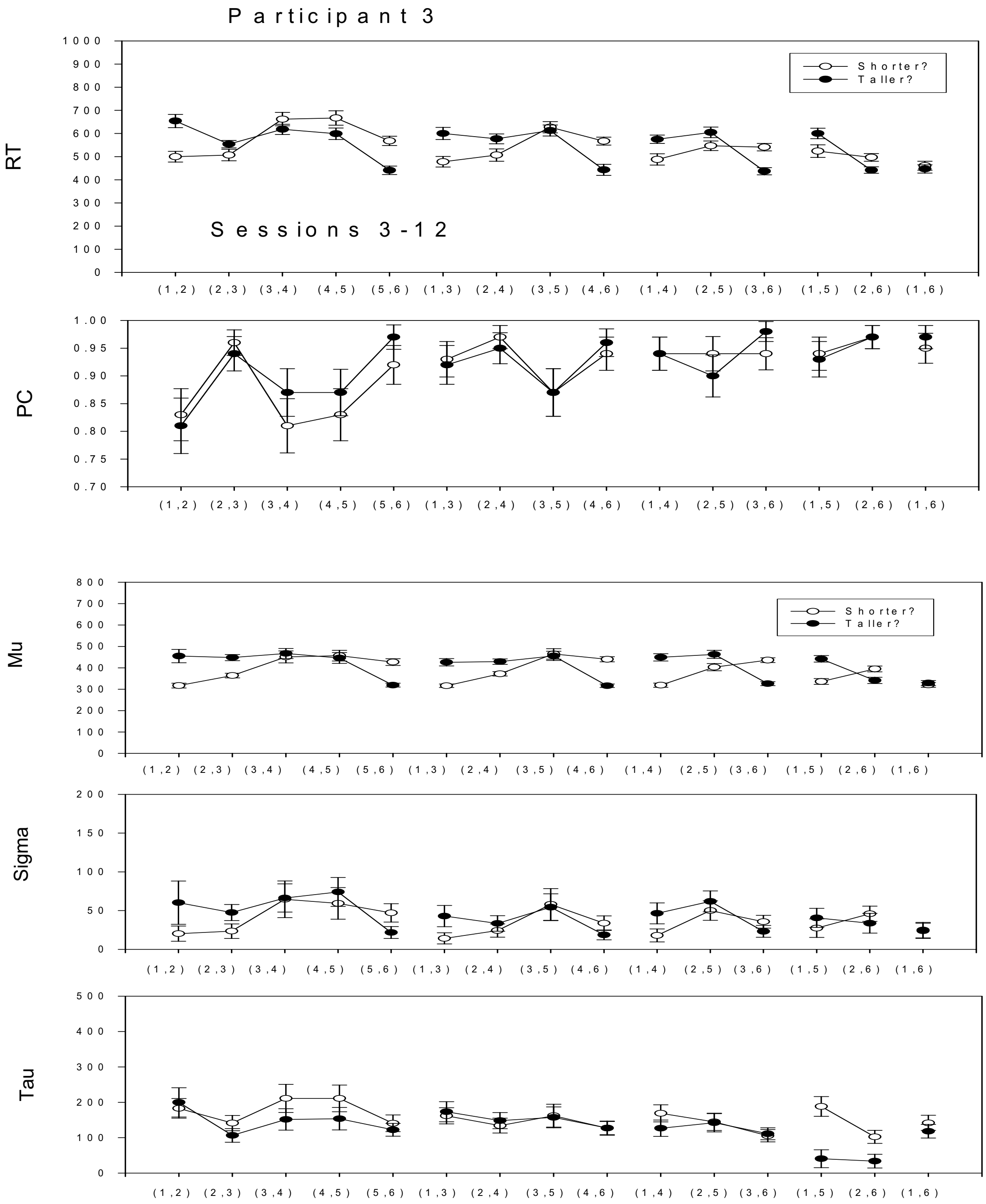
Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446-461.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190-212.


Participant 1
Sessions 3-12
RT
PC
Shorter?
Taller?
(1,2) (2,3) (3,4) (4,5) (5,6) (1,3) (2,4) (3,5) (4,6) (1,4) (2,5) (3,6) (1,5) (2,6) (1,6)


Mu
Sigma
Tau
Shorter?
Taller?
(1,2) (2,3) (3,4) (4,5) (5,6) (1,3) (2,4) (3,5) (4,6) (1,4) (2,5) (3,6) (1,5) (2,6) (1,6)


Participant 2
Sessions 3-12
Shorter?
Taller?
RT
PC
Mu
Sigma
Tau
(1,2) (2,3) (3,4) (4,5) (5,6) (1,3) (2,4) (3,5) (4,6) (1,4) (2,5) (3,6) (1,5) (2,6) (1,6)


Participant 3
Sessions 3-12
Shorter?
Taller?
RT
PC
Mu
Sigma
Tau
(1,2) (2,3) (3,4) (4,5) (5,6) (1,3) (2,4) (3,5) (4,6) (1,4) (2,5) (3,6) (1,5) (2,6) (1,6)

CHANNEL-INDEPENDENT AND SENSOR-INDEPENDENT STIMULUS REPRESENTATIONS

David N. Levin
Department of Radiology, University of Chicago, Chicago, IL 60637 (d-levin@uchicago.edu)

Abstract

This paper shows how to design intelligent sensory devices that represent stimulus relationships in a channel-independent and sensor-independent manner. First, we demonstrate that a device's channel and sensor define a coordinate system that the device imposes on the space of stimulus states. Then, differential geometry is applied to derive coordinate-system-independent stimulus relationships from statistical characteristics of the time series of previously-encountered sensor measurements. This procedure is illustrated with analytic examples. In an intelligent sensory device, this kind of representation "engine" could function as a "front end" that passes channel/sensor-independent stimulus representations to a pattern recognition module. After a pattern recognizer has been trained in one of these devices, it could be used without change in other devices having different channels and sensors.

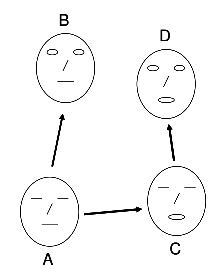


Figure 1. Consider a two dimensional space populated by faces characterized by the extent to which the mouth and eyes are open. Given two faces (*A* and *B*) that are related by a small transformation ($A \to B$, called an "eye-opening" transformation) and given another face (*C*) that differs from *A* by another small transformation ($A \to C$, called a "mouth-opening" transformation), observers will often judge a particular transformation of *C*, namely $C \to D$, to be perceptually equivalent to $A \to B$. In other words, they will perceive the stimuli as being related by the analogy: *A:B = C:D*.

A method of designing channel-independent and sensor-independent sensory devices is suggested by the following discussion of human perception. People usually describe their perceptions in a relative manner; i.e., in relation to their perceptions of other stimuli. Thus, if an individual is asked to describe a light or a shape or a sound, he/she is likely to compare it to his/her perception of other lights or shapes or sounds. People may make certain "primitive" perceptual judgments that are used to construct the perceived relationships between stimuli. For example, in most circumstances, observers have a strong sense about whether two stimuli are the same or not; i.e., whether stimulus $A = B$ or not. Likewise, observers often feel that stimulus "analogies" are true or false. For instance (Fig. 1), suppose that stimuli *A*, *B*, *C*, and *D* differ by small stimulus transformations. Observers will often have a definite sense about whether the relationship between *A* and *B* is the same (or is not the same) as the relationship between *C* and *D*; i.e., whether *A:B = C:D* or not. A series of such perceived analogies can be concatenated in order to describe relationships between two "distant" stimuli (Levin, 2000). Figure 2 shows an example in which an observer describes stimulus *E* as being related to stimuli *A*, *B*, and C by a succession of analogous stimulus transformations: "*E* is the stimulus that is produced by starting with stimulus *A* and performing three transformations perceptually equivalent to the $A \to C$ transformation, followed by two transformations perceptually equivalent to the $A \to B$ transformation". In principle, an observer, who perceives local stimulus analogies at each point in stimulus space, can use this technique to describe the paths connecting any two stimuli of interest. This knowledge of relative stimulus locations can be used to navigate through stimulus space, and it comprises a kind of representation of the stimulus space (a kind of "map of the world") in the same sense that a geographic representation of a city is provided by the knowledge of how to navigate between any two points in it.

If the brain uses stimulus analogies to impose order on the "world" of stimuli, it seems likely that these analogies are derived from past sensory history, rather than being "hard-wired" from birth. For example, there is evidence that young children "learn to see", and adults with no previous visual experience usually do not sense meaningful relationships between visual stimuli. The latter phenomenon is illustrated by many neurological case histories in which congenital cataracts were excised from the eyes of a blind adult with the intention of allowing the patient to see for the first time. In each case, when the post-operative bandages were removed, the patient invariably experienced an inchoate mixture of meaningless visual patterns, presumably because he/she never learned how to impose order on visual information. Normal individuals with similar past experiences learn to perceive the world in approximately the same way, and this is remarkable, given the fact that different individuals have different "sensors" (e.g., different eyes, ears, and primary sensory cortices). To illustrate, suppose that stimuli *A*, *B*, *C*, and *D* create complex spatial distributions of neuronal electrical discharges (denoted x_A, x_B, x_C, and x_D) in the primary sensory cortex of observer Ob. The same stimuli will produce different three-dimensional electrical current distributions (x'_A, x'_B, x'_C, and x'_D) in the sensory cortex of a second observer Ob', who has a different sensory system Yet, despite these differences, if the two observers have had similar life histories and Ob senses $x_A : x_B = x_C : x_D$, then Ob' tends to independently conclude $x'_A : x'_B = x'_C : x'_D$. A related fact is that each individual's perceptions of stimuli tend to be independent of the channel through which he/she views those stimuli. This fact was strikingly illustrated by experiments in which subjects wore goggles that created severe geometric distortions and inversions of the observed scenes (Held, 1972). Although each subject initially perceived the distortion of the new channel, his/her perceptions of the world tended to return to the pre-experimental baseline after several weeks of constant exposure to familiar stimuli seen through the goggles. These observations suggest the following generalization. Consider an individual who has a certain history of stimulus exposure and "learns" to perceive relationships (e.g., $x_A : x_B = x_C : x_D$) among certain stimulus-induced sensory cortex states. If that person is exposed to a similar history of stimuli through a new observational channel, he/she will relearn relationships among sensory cortex states with the result that the altered sensory cortex states induced by stimuli through the new channel will be perceived to be related to one another in the same way (e.g., $x'_A : x'_B = x'_C : x'_D$) as the corresponding unaltered cortex states were related before the channel changed. *In summary: the phenomena discussed in this Section suggest that the brain uses its past experience to derive stimulus relationships (e.g., stimulus analogies and relative stimulus locations) that are invariant in the presence of channel-induced and sensor-induced transformations of sensory cortex states.*

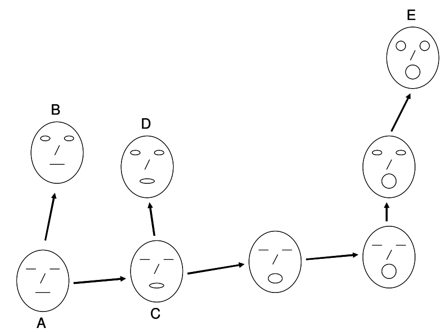


Figure 2. "Distant" stimuli can be related by describing a path connecting them as a sequence of perceptually defined stimulus transformations. For example, an observer may describe *E* as the stimulus that is produced by starting at *A* and performing three transformations perceptually equivalent to the mouth-opening transformation $A \rightarrow C$, followed by two transformations perceptually equivalent to the eye-opening transformation $A \rightarrow B$.

In this paper, we propose how to design sensory machines that also represent relative stimulus locations in a channel-independent and sensor-independent manner (Levin, 2001, 2004c). Consider a sensory device having the general architecture shown in Figure 3, and suppose that the device is exposed to a "world" consisting of an evolving stimulus. Energy from the stimulus traverses a channel, before being detected and processed by a sensor. This processing defines a sensor state (*x*), which is a function of the stimulus state and is analogous to the state of the brain's primary sensory cortex. In a conventional sensory device, the sensor state would be sent directly to a pattern recognition module, and, if the channel or sensor changes, the device must be taken "off-line" for retraining of the recognizer or recalibration of the detector. In the devices proposed in this paper, the

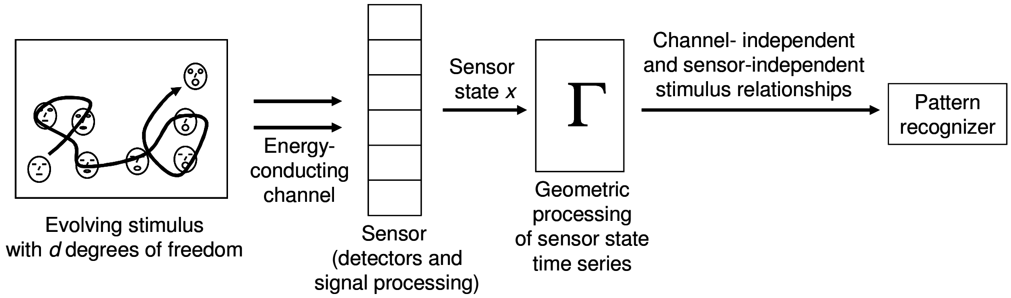


Figure 3. The architecture of an intelligent sensory device. The energy from the stimulus traverses a channel, before being detected and processed by a sensor. The resulting collection of measurements is used to define a sensor state. The time series of sensor states is processed in order to find channel-independent and sensor-independent relationships among stimulus states, which comprise the input of a pattern recognizer.

sensor state time series is first processed in order to learn how to represent each stimulus state by its channel-independent and sensor-independent location relative to previously-encountered stimulus states. Because pattern recognition is applied to this information, it is not channel-dependent or sensor-dependent.

Previously-reported methods of data representation do not have the desired channel-independence and sensor-independence. For example, in multidimensional scaling (Cox, 1994) and Isomap (Tenenbaum, 2000), the representation of stimulus states depends on the assignment of distances between each pair of sensor states, and these distances change if the stimulus measurements are subjected to a nonlinear transformation, caused by a change of channel or sensors. Likewise, in locally-linear embedding (Roweis, 2000), stimulus states are represented by coordinates that are not invariant under all of the stimulus measurement transformations, which are induced by channel and sensor changes. And even less powerful methods of data representation, such as principal components analysis, have the same drawback.

In the next section, we show that the problem of finding channel-independent and sensor-independent relative stimulus locations can be mapped onto the differential geometric problem of finding coordinate-system-independent properties of the trajectory of previously-encountered stimulus configurations. In Examples, we describe large classes of stimulus trajectories for which the desired invariant stimulus representations can be analytically derived. The implications and applications of this approach are discussed in the last section.

Theory

Suppose that the stimulus is some physical system that has d degrees of freedom, and suppose that the observing machine's sensor makes at least $2d+1$ measurements, which are time-independent functions of the stimulus configuration. This means that the channel and sensor are assumed to be stationary, although this assumption can be weakened if the device is run in an adaptive mode (Discussion section). The measurement functions define a mapping from the d-dimensional stimulus space to a d-dimensional subspace of the higher-dimensional ($\geq 2d+1$) measurement space. The Takens embedding theorem, which is well-known in the field of nonlinear dynamics (Sauer, 1991), states that this mapping is invertible for almost all choices of measurement functions. Essentially, invertibility is likely because so much "room" is provided by the "extra" dimensions of the higher dimensional space that the d-dimensional subspace, which is the range of the mapping, is very unlikely to self-intersect. Suppose that the stimulus evolves so that the stimulus trajectory densely covers a patch of stimulus space and the corresponding sensor measurements densely cover a patch of the d-dimensional measurement subspace. Then, the machine can use dimensional reduction methods (Roweis, 2000; Tenenbaum, 2000) to learn the location and shape of the measurement subspace, and it can impose an arbitrarily-chosen x (x_k , $k = 1, 2, \ldots, d$) coordinate system on it. The quantity x is defined to be the sensor state of the machine, and, according to the embedding theorem, it is invertibly related to the stimulus state. For example, suppose that the stimulus consists of a particle that is moving on a lumpy transparent two-dimensional surface in a laboratory ($d = 2$). Furthermore, suppose that its state is being monitored by three video cameras in various locations and positions, each of which derives two measurements from the imaged scene (e.g., the two pixel coordinates of the particle or two Fourier coefficients of the image or two wavelet coefficients of the image or two …). Notice that the mapping between the particle's position and the measurements of any one camera may not be invertible because of the lumpiness of the surface. As the particle moves over the surface in the laboratory, the six camera measurements move in a two-dimensional subspace within the six-dimensional space of possible measurements, and the location and shape of this subspace can be learned by dimensional reduction of a sufficiently dense set of measurements. Each stimulus configuration (each position of the particle in the laboratory) produces a sensor state x, which is defined to be the location of the corresponding measurements in some two-dimensional coordinate system on the measurement subspace. The embedding theorem asserts that the mapping between the stimulus and sensor states is one-to-one, for almost all choices of camera positions and orientations and for almost all types of image-derived measurements.

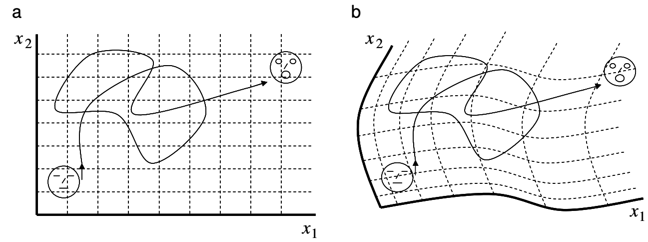


Figure 4. Because of the embedding theorem, devices with different channels and/or sensors can be considered to be viewing the same stimulus trajectory in different coordinate systems of the stimulus state space (panels *a* and *b*).

In general, our objective is to represent the locations of stimulus states relative to one another in a channel-independent and sensor-independent manner. The introductory discussion suggests that the machine should learn these relationships from its "life experience", i.e., from the sequence of previously-observed stimuli. In any event, the trajectory of previously-encountered stimuli provides the only possible source of such "structure" on the stimulus space. Because the mapping between the stimulus and sensor states is invertible, the sensor state can be considered to be the location of the stimulus in a particular (x) coordinate system on the stimulus space. Therefore, when machines with different channels and sensors observe the same stimulus time series, they are simply using different coordinate systems to view the same trajectory in the stimulus space (Fig. 4). Hence, the problem of defining channel-independent and sensor-independent relative stimulus locations is the same as the problem of finding coordinate-system-independent stimulus locations. Recall that Euclidean geometry provides methods of making statements about a geometric figure that are true in all rotated or translated coordinate systems. Similarly, differential geometry provides the mathematical machinery for making statements about a geometric figure (such as the stimulus trajectory) that are true in all non-linearly transformed coordinate systems (Weinberg, 1972). In other words, differential geometry seeks to find properties of a geometric figure that are "intrinsic" to it in the sense that they don't depend on extrinsic factors, such as the coordinate system in which the observer chooses to "view" it (i.e., the choice of numerical labels assigned to the figure's points). This suggests the following strategy: 1) first, use differential geometry to find coordinate-system-independent properties of the trajectory of previously-encountered stimuli; 2) then, use those properties to derive coordinate-system-independent stimulus analogies at each point of the traversed part of the stimulus space; 3) finally, use these analogies to derive relative stimulus locations (e.g., Fig. 2) that are coordinate-system-independent (and, therefore, channel-independent and sensor-independent).

From a geometric point of view, a local stimulus analogy specifies a coordinate-system-independent way of moving one short line segment along a second short line segment in order to create a third short line segment that is said to be "equivalent" to the first one. For example, in Fig. 1, $C \to D$ is the line segment at C that is perceptually equivalent to the line segment $A \to B$ at A, after it has been moved along the line segment $A \to C$. This is exactly the operation that is performed by the parallel transfer procedure of differential geometry, which comprises a coordinate-system-independent way of designating vectors along a path that are said to be equivalent to a given vector at the path's origin (Fig. 5). Therefore, if a parallel transfer operation can be derived from the stimulus trajectory, it can be used to define stimulus analogies, and these can be concatenated to define relative stimulus locations in a coordinate-system-independent manner. There are a few methods of using the stimulus trajectory to define a parallel transfer operation (Levin, 2001, 2002a, 2004a), and in this paper, we focus on one of them. First, we use the stimulus trajectory to derive a metric on stimulus space (i.e., a coordinate-system-independent method of defining the lengths of line segments and other vectors). Then, we derive a parallel transfer mechanism from that metric by using the following fact: there is a unique parallel transfer procedure that preserves the lengths of transferred vectors with respect to a given metric without producing torsion (without producing reentrant geodesics, [Weinberg, 1972]). In any coordinate system (x), this particular parallel transfer operation can be computed in the following manner (Fig. 5): given the metric tensor $g_{kl}(x)$, a vector V at x is parallel transferred along a line segment δx into the vector $V + \delta V$ at $x + \delta x$, where

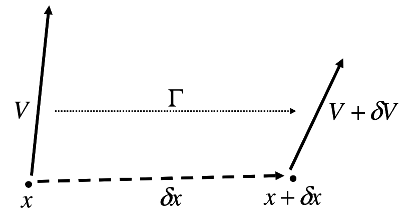


Figure 5. The parallel transfer procedure Γ moves a vector V at x along the short path segment δx to produce an "equivalent" vector $V + \delta V$ at $x + \delta x$.

$$\delta V^k = - \sum_{l,m=1,\ldots,d} \Gamma^k_{lm}(x) V^l \delta x_m \,, \tag{1}$$

where $\Gamma^k_{lm}(x)$ (the affine connection at x) is given by

$$\Gamma^k_{lm} = \frac{1}{2} \sum_{n=1,\ldots,d} g^{kn} \left(\frac{\partial g_{mn}}{\partial x_l} + \frac{\partial g_{nl}}{\partial x_m} - \frac{\partial g_{lm}}{\partial x_n} \right) \tag{2}$$

and where g^{kl} is the inverse of g_{kl}. Therefore, the whole problem of deriving channel-independent and sensor-independent stimulus relationships has been reduced to the task of using the stimulus trajectory to derive a non-singular metric (i.e., a non-singular second rank symmetric covariant tensor). Once that metric has been found, Eqs. 1-2 can be used to define a parallel transfer operation, which makes it possible to define stimulus analogies and relative stimulus locations in a coordinate-system-independent manner (and, therefore, in a channel-independent and sensor-independent manner).

Let $x(t)$ describe the evolving stimulus state in the coordinate system defined by a particular observing machine's sensor. At each point y, consider the local covariance matrix of the trajectory's "velocity"

$$c^{kl}(y) = \langle \dot{x}_k \dot{x}_l \rangle_{x(t) \sim y} \tag{3}$$

where $\dot{x} = dx/dt$ and the bracket denotes the time average over the trajectory's segments in a small neighborhood of y. If this quantity approaches a definite limit as the neighborhood shrinks around y, it will certainly transform as a symmetric contravariant tensor. Furthermore, as long as the trajectory has d linearly-independent segments passing near y, this tensor can be inverted to form a symmetric covariant tensor, $g_{kl}(y) = (c^{-1})_{kl}$. Thus, Eq. 3 can be used to derive a metric from any stimulus trajectory for which the above-described limit exists.

Examples

In this Section, we demonstrate large classes of stimulus trajectories for which the local velocity covariance matrix (and metric) are well-defined and can be computed analytically. Specifically, we construct these trajectories from the behavior of physical systems that can be realized in the laboratory. First, consider a system with d degrees of freedom, and suppose that the system's energy is given by

$$E(x,\dot{x}) = \frac{1}{2}\sum_{k,l=1,\ldots,d} \mu_{kl}(x)\dot{x}_k\dot{x}_l + V(x) \tag{4}$$

where μ_{kl} and V are some functions and x is the stimulus' location in some coordinate system. Equation 4 is the energy function of a large variety of physical systems in which the degrees of freedom can be spatial and/or internal and in which d may be large or small. The simplest system with this energy is a single particle with unit mass that is moving in potential *V(x)* on a possibly-curved two-dimensional frictionless surface with physical metric μ_{kl}. For example, if the particle is moving on a frictionless spherical, cylindrical, or planar surface in the laboratory, μ_{kl} is the metric induced on the surface by the Euclidean metric in the laboratory's coordinate system. Returning to the more general case, suppose that the system intermittently exchanges energy with a thermal "bath" at temperature *T*. In other words, suppose that the system evolves along one trajectory from the Boltzmann distribution at that temperature and periodically jumps to another randomly-chosen trajectory from that distribution. After a sufficient number of jumps, the amount of time the particle will have spent in a small neighborhood dx of x and a small neighborhood $d\dot{x}$ of $\dot{x}$ is proportional to the Boltzmann distribution (Reif, 1965)

$$\mu(x)\exp\left[-E(x,\dot{x})/kT\right]dx\,d\dot{x} \tag{5}$$

where k is the Boltzmann constant and μ is the determinant of μ_{kl}. If this expression is substituted into the right side of Eq. 3 and the resulting Gaussian integrals are evaluated, the velocity covariance matrix is found to be well-defined and given by

$$c^{kl}(x) = kT\mu^{kl}(x) \tag{6}$$

where μ^{kl} is a contravariant tensor equal to the inverse of μ_{kl}. It follows that the trajectory-induced metric on the stimulus space is $g_{kl}(x) = \mu_{kl}(x)/kT$; i.e., it is proportional to the physical metric on the surface. Thus, a metric is induced on the stimulus state space by the trajectories of each system in this large class of systems. It follows that Eqs. 1-2 define a coordinate-system-independent method of moving line segments across the stimulus manifold, and this procedure can be used to define channel-independent and sensor-independent stimulus analogies and relative stimulus locations. As an illustration, reconsider the case (Theory section) of the particle that is moving on a surface and is being observed by a machine equipped with multiple cameras. Two observing machines of this type will agree on all statements about relative particle locations (i.e., all statements about parallel transfer operations, like those in Fig. 2), even though their sensors and channels may be dramatically different (e.g., different numbers of cameras, different camera positions and orientations, camera lenses producing different distortions of the observed scene, different image-derived measurements such as particle pixel coordinates vs. image Fourier coefficients vs. …).

In this Section, we demonstrated large classes of trajectories that endow the stimulus manifold with a metric (and, consequently, channel-independent and sensor-independent stimulus relationships). It is likely that metrics can be derived from many other trajectories, especially those with local velocity distributions that vary sufficiently smoothly across stimulus space. However, it is an experimental question whether this is the case for the trajectories of specific physical stimuli. For example, it has been shown that a metric can be derived from the stimulus trajectories produced by spoken English. In these experiments, the metric was used to derive a channel-independent representation of speech trajectories, with the ultimate goal of achieving channel-independent automatic speech recognition. A complete description of these experiments is available in Levin (2004a).

Discussion

We have shown how a machine, which observes stimuli through a specific channel and with specific sensors, can glean channel-independent and sensor-independent information about the locations of stimuli relative to one another. This is possible if two requirements are satisfied: 1) the observed stimulus trajectory has a well-defined local velocity covariance matrix; 2) the device measures more than *2d* properties of the stimulus, where *d* is the number of stimulus degrees of freedom. Like a human, such a machine represents the world in a way that depends on the statistical properties of its past sensory "experience". Two such machines, using different channels and sensors, will represent stimulus relationships in the same way, as long as their past trajectories through stimulus space impose the same geometry on that space, and this will be true as long as the velocity covariance matrices of their trajectories are identical when expressed in the same stimulus coordinate system. On the other hand, if their stimulus trajectories have different velocity covariance matrices, the two machines will not agree on all statements about relative stimulus locations. For example, there will be striking disagreements if the trajectory-derived metric of one machine is flat and the trajectory-derived metric of the other one is curved (Levin, 2000). The methodology in this paper should not be confused with conventional blind source estimation techniques. The latter attempt to estimate the form of the stimulus trajectory when it is measured in a specific manner (e.g., in a specific coordinate system). In this paper, we seek to determine properties of the stimulus trajectory that are independent of the way it is measured (i.e., coordinate-system-independent).

It is interesting to consider the kinds of "memories" that one of these devices could have. A neural network or parametric method could be used to store the location and shape of the *d*-dimensional measurement subspace, once the machine has learned that information by observing a sufficiently dense stimulus trajectory. This subspace contains all sensor state measurements encountered in the "world". It could also store the affine connection, which encodes the relative locations of sensor states in that subspace. For example, the weights of a neural network could be adjusted so that it comprises a kind of "connection engine" for mapping each sensor state x onto $\Gamma^k_{lm}(x)$. If the machine subsequently re-experiences a sensor state (e.g., the sensor state corresponding to stimulus *E* in Fig. 2), it can use the stored affine connection to deduce it's location relative to other nearby sensor states in the measurement subspace (e.g., *A*, *B*, and *C* in Fig. 2). Thus, previously-experienced stimulus relationships can be recomputed at any time, as long as the machine's memory contains the previously-experienced affine connection. On the other hand, suppose that the parallel transfer mechanism has changed since a stimulus was first encountered, because the machine has experienced a statistically different stimulus trajectory in the intervening time. Then, the machine will describe the relative location of a stimulus differently than it did before. Likewise, the trajectory segment of a recalled train of events may be described differently with the new affine connection than it was with the affine connection that existed at the time those events first transpired. Despite these differences, each description of the events is valid at the time it was derived, in the sense that it is coordinate-system-independent (and, therefore, independent of factors extrinsic to the stimulus trajectory, such as the nature of the machine's channel and sensors).

The nonlinear signal processing method presented in this paper could be used as a representation "engine" in the "front end" of an intelligent sensory device (Levin, 2004b). It would produce channel-independent and sensor-independent stimulus locations that could be used by the device's pattern recognition module in order to recognize stimuli. Because the effects of channels and sensors have been "filtered out" of these representations, devices with different channels and sensors could use the same pattern recognition module, without recalibrating their detectors or retraining the recognizer. As described in Levin (2004c), this may make it possible to design detector-independent computer vision devices. Furthermore, because the method does not explicitly depend on the nature of the detectors, it is a natural way to achieve multimodal fusion of audio and optical sensors. Finally, there is the possibility of creating a speech-like telecommunications system that is resistant to channel-induced corruption of the transmitted information (Levin, 2004a). In such a system, information is carried by the coordinate-system-independent relationships among the transmitted power spectra.

Up to now, we have assumed that the channel and sensor of a machine are time-independent and that a machine derives a parallel transfer mechanism from the history of all previously-encountered stimuli. However, if a machine derived its affine connection from stimuli encountered in the most recent time interval ΔT, it could adapt to channels and/or sensors that drift over a longer time scale. It would continue to represent relative stimulus locations in the same way, as long as the velocity covariance matrices of the stimulus trajectory were temporally stable. Notice that the trajectories constructed in Examples have this kind of temporal stability. This process of adaptation, which is analogous to human adaptation in the goggle experiments, was demonstrated with simple experimental examples in Levin (2002b).

The remarkable channel-independence and sensor-independence of human perception have been the subject of philosophical discussion since the time of Plato (e.g., the allegory of "The Cave"), and these issues have also intrigued modern neuroscientists. How can an individual perceive the intrinsic constancy of a stimulus even though its appearance is varying because of changing observational conditions? Why do different people tend to have similar perceptions despite significant differences in their sensory organs and neural processing pathways? This paper shows how to design sensory devices that invariantly represent stimuli in the presence of processes that transform their sensor states. These stimulus representations are invariant because they encode "inner" properties of the time series of stimulus configurations themselves; i.e., properties that are largely independent of the way the observer views them (i.e., independent of the choice of sensor and channel). Significant evolutionary advantages would accrue to organisms that developed the ability to process sensory information in this way. For instance, they would not be confused by the appearance-changing effects of different observational conditions, and communication among different organisms would be facilitated because they independently represent stimuli in the same way. Biological experimentation is required to determine whether humans and other species achieve these objectives by means of the general approach in this paper. However, it would not be surprising if they did, because there are not many other ways of representing the world in an invariant fashion.

References

Cox, T. and Cox, M. (1994). *Multidimensional Scaling.* London : Chapman & Hall.

Held, R. and Whitman, R. (1972). *Perception: Mechanisms and Models.* San Francisco: Freeman.

Levin, D. N. (2000). A differential geometric description of the relationships among perceptions. *J. Math. Psych., 44*, 241-284.

Levin, D. N. (2001). Stimulus representations that are invariant under invertible transformations of sensor data. *Proc. SPIE, 4322*, 1677-1688.

Levin, D. N. (2002a). Sensor-independent stimulus representations. *Proc. Nat. Acad. Sci. (USA), 99,* 7346-7351.

Levin, D. N. (2002b). Representations of sound that are insensitive to spectral filtering and parameterization procedures. *J. Acoust. Soc. Am., 111*, 2257-2271.

Levin, D. N. (2004a). Blind normalization of speech from different channels. *J. Acoust. Soc. Am., 116,* 519-529.

Levin, D. N. (2004b). US Patent #6,687,657 and pending patents.

Levin, D. N. (2004c). See papers at http://www,geocities.com/dlevin2001/

Reif, F. (1965). *Fundamentals of Statistical and Thermal Physics.* New York : McGraw-Hill.

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science, 290*, 2323-2326.

Sauer, T., Yorke, J. A., and Casdagli, M. (1991). Embedology. *J. Stat. Phys., 65*, 579-616.

Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science, 290*, 2319-2323.

Weinberg, S. (1972). *Gravitation and Cosmology.* New York: John Wiley & Sons.

THE STRUCTURE OF CONTEXT

Stephen Link
University of California, San Diego
link@ucsd.edu

Abstract

Context is the cognitive milieu that the mind creates for itself. This paper investigates effects of context on judgment and decision making and suggests a method for determining an underlying structure for linguistic context.

The Oxford English Dictionary provides four definitions of the word "context." First, as "the weaving together of words and sentences; construction of speech, literary composition." Second, "The connected structure of a writing or composition." Third, "The connection or coherence between the parts of a discourse. Fourth, "The whole structure of a connected passage regarded in its bearing upon any of the parts that constitute it." Somehow the idea that a mind must create the "weaving together," the connected structure," the coherence," and how the whole structure of a connected passage bears on the parts that constitute it seems to have escaped the notice of our erudite scholars. Let me add a psychophysicist's definition. The word "context" refers to the cognitive milieu in which the mind finds itself.

Context in experimental designs

As psychophysicists we can create methods for measuring mental phenomena such as "context." As a start, Helson's Adaptation Level provides a measure of the effect of the experimental context on the basis for discriminative judgments. Parducci's Range-Frequency Theory shows how context created by the range of sensory stimulation and the frequency of stimuli affect discriminative judgments. I was very much influenced by Parducci's theory in creating the Method of Symmetric Differences (MSD) which provides a constant range of comparison stimuli while varying the size of a standard stimulus in the middle of this range - a context for judgments.

The theoretical reasons for creating MSD derive from problems generated by the Method of Paired Comparisons (MPC). In classic MPC, developed by Thurstone (1927), all stimuli are to be compared against each other. Thurstone's application to the study of social values required each of 266 University of Chicago students to compare one of 19 crimes against another on the basis of seriousness. Using all $n = 19$ crimes yielded a total of $(n)*(n-1)/2 = 171$ unique paired comparisons. Each student judged each pair of crimes only once. From the frequencies of choice Thurstone derived a measure of seriousness of each crime separated from another. From these measures Thurstone derived a scale of the seriousness of crimes. The experiment and method were used again by Coombs with students from the University of Michigan and later by Link with students from McMaster in Ontario, Canada. In both these replications the scale for the seriousness of crime remained comparable to those obtained by Thurstone in 1927. The judgments are contextual in the sense that the population of judges is university students. Judgments made by other groups, or in other societies, may expose important cultural differences.

MPC is itself a context with implications for the meaningfulness of the resulting scale values. Guilford (1932), in an effort to provide a psychological scale for the heaviness of weights, used as comparison stimuli weights raging to 215g from 185g in equal steps of 5g. In this application of the method a subject made 100 judgments of each possible pair in order to provide the frequencies of response needed to apply Thurstone's theory. During the course of this lengthy experiment subjects easily learn the range of weights used by the experimenter. Because of the nature of the judgment process one weight must be sensed before another. Then, when presented an extremely light weight as a standard against which a second weight must be judged, the subject can be almost certain that the second comparison weight will feel heavier. The subject may then respond on the basis of this preparatory knowledge without using the sensory information provided by lifting the second weight. This response bias may even vary across the range of the standards. A very heavy weight presented as a standard may induce a bias toward responding "lighter" to the comparison weight. But, in the original theory, there is no measure of this effect of the experimental context on discriminative judgments.

The Method of Paired Comparisons (MPC) creates a context that induces response bias. The effects of response bias are particularly evident in response times. For example, The Method of Serial Comparisons (Link, 1990a) was used to generate comparative judgments between two digit numbers presented sequentially. The number 55 is presented as a starting value. The subjects are to judge whether the next number, selected at random from digits ranging to 99 from 11 (sans 55), is larger or smaller. This number is used as the basis for the comparison of the third randomly selected number, and so forth. When subjects are presented an extreme value they can be quite certain that the following number will be less extreme. This response bias leads to very fast responses at the extremes of the range of stimuli.

The Method of Symmetric Differences combats this bias (Link, 1990b). The purpose is to ensure that subjects cannot use the magnitude of a standard stimulus as a basis for response bias. Thus, comparison stimuli larger or smaller than the standard are presented at random with equal frequency. A standard of 33 could be followed by a comparison number ranging to 55 from 11 with equal frequency. Numbers from 33 to 77 could be compared against a standard of 55. Numbers from 55 to 99 could be compared against a standard of 77. In all cases 22 comparison stimuli are greater than the standard and 22 are less. An example of the design matrix appears in Figure 1.

		STIMULUS DIFFERENCE									
		-22	-21	-20	…	-1	1	…	20	21	22
	77	55	56	55	…	76	78	…	97	98	99
	76	54	55	56	…	75	77	…	96	97	98
	75	53	54	55	…	74	76	…	95	96	97
	…	…	…	…	…	…	…	…	…	…	…
STANDARD	56	34			…	55	57	…			76
STIMULUS	55	33			…	54	56	…			75
	54	32			…	53	55	…			74
	…	…	…	…	…	…	…	…	…	…	…
	35	13			…	34	36	…			57
	34	12			…	33	35	…			56
	33	11	12	13	…	32	34	…	53	54	55

Figure 1. Entries are values for stimuli compared against a standard shown in the left-most column. For some experiments, such as numerical comparisons, the 0 stimulus difference values are not used although in other experiments they provide a basis for a measure of response bias.

Two distinctive contextual features of MSD are a uniform probability distribution over values of standards and stimulus difference (ΔS) and a triangular probability distribution over values of comparison stimuli. In MPC the uniform distribution is over standards and comparison stimuli and the triangular distribution is over stimulus difference (ΔS). In the MPC the presentation of a standard equal to 33 informs the subject that the number presented for comparison is likely to be larger than 33. In MSD the presentation of any standard provides no information about the likelihood of a response.

The design matrix in Figure 1 contains a total of 1936 cells. The value 55 was used as a standard in practice trials and comparison values of 55 ± 22 were randomly selected for these 44 trials. No values for 0ΔS were used. Time limitations of one hour for an experimental session placed severe restrictions on the number of trials per session. This constraint led to creating four experimental sessions of ¼(1936-44=1896) = 462 randomly selected experimental trials, preceded by 11 practice trials. Thus, four one hour sessions were required to complete all presentations shown in the design matrix.

Let's examine MPC in the context of an experiment on the comparison of dot distances. The stimuli are two white dots horizontally displaced by a number of millimeters ranging to 99 from 11, and appear about 1m distant from the subject on a compute-controlled screen with a black background. A standard, ranging to 77mm from 33mm, is presented at the beginning of a self-paced trial for 200 milliseconds. Then, after a 200 msec dead time, with no display, the comparison pair of dots is presented until a response occurs. The horizontal distance between this pair will always be of greater or lesser distance than the standard on half the trials, and range from -22mm to +22mm. For example, a standard distance of 44mm will be followed by a comparison stimulus ranging from 22mm to 66mm (sans 44mm).

Subjects responded by releasing a trial initiation key (TIK) and then depressing one of six response buttons arranged in a semi-circle about 8cm distant from TIK. The six response keys were assigned confidence values of HIGH, MEDIUM, LOW, LOW, MEDIUM, HIGH from left to right. For some subjects the 3 response keys on the left indicated the response "SMALLER" and the three right-hand keys "LARGER." For other subjects the response "LARGER" was assigned to the leftmost response keys and the response "SMALLER" to the rightmost keys.

Each of five subjects completed a replication of the design matrix for practice, and two more experimental replications for a total of 18,480 trials. For the purpose of determining a numerical measure of confidence, values of -5, -3, -1, 1, 3, 5 were assigned as choice-confidence values for the responses SMALLER-HIGH, SMALLER-MEDIUM, SMALLER-LOW, LARGER-LOW, LARGER-MEDIUM, and LARGER-HIGH. For each response, the time taken from the onset of the comparison stimulus to the depression of the response key measured total response time.

Figure 2 illustrates mean response time on the right-hand ordinate, average proportion of "LARGER" responses, and mean response confidence as a function of stimulus difference on the left-hand ordinate. The psychometric "LARGER" function ranged to 0.99 from 0.01 and shows a steady increase, even in this constrained scale, across stimulus differences ranging to 22 from -22. The mean response confidence changes steadily from -4.7 to +4.7 as stimulus differences increase. Mean response time is maximized for the smallest values of stimulus difference. Although this chronometric function shows a reduction in RT for very negative stimulus differences, larger stimulus differences show a somewhat greater mean response time, resulting in a slight lack of symmetry.

These results are quite comparable to those obtained using the Method of Constant Stimuli with a single standard. However, these average results yield a displacement of the point of subjective equality, where the proportion of "LARGER" responses equals 0.50, to the right of zero stimulus

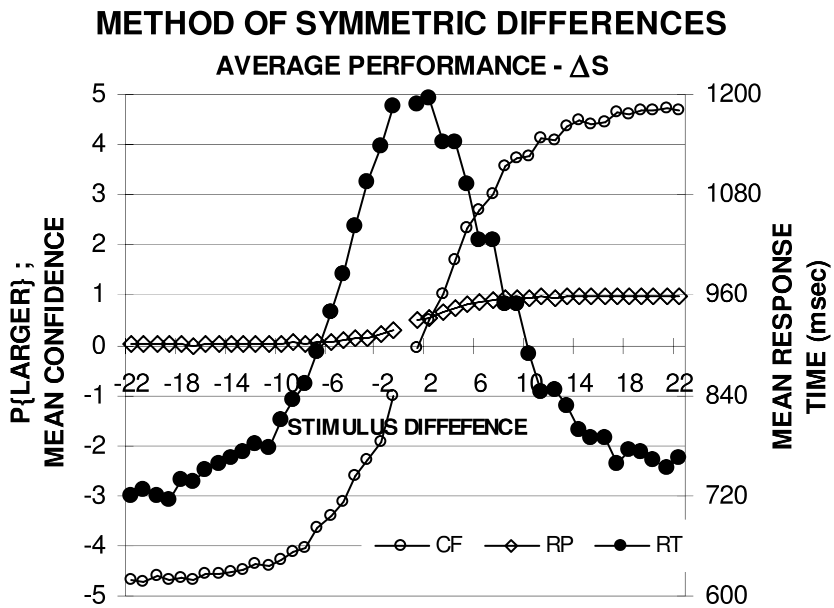


Figure 2. Average performance for 5 subjects for a total of 18,480 trials. Each stimulus difference occurs 420 times. The labels are CF (Mean confidence), RP (Larger response proportion) and RT (Mean response time).

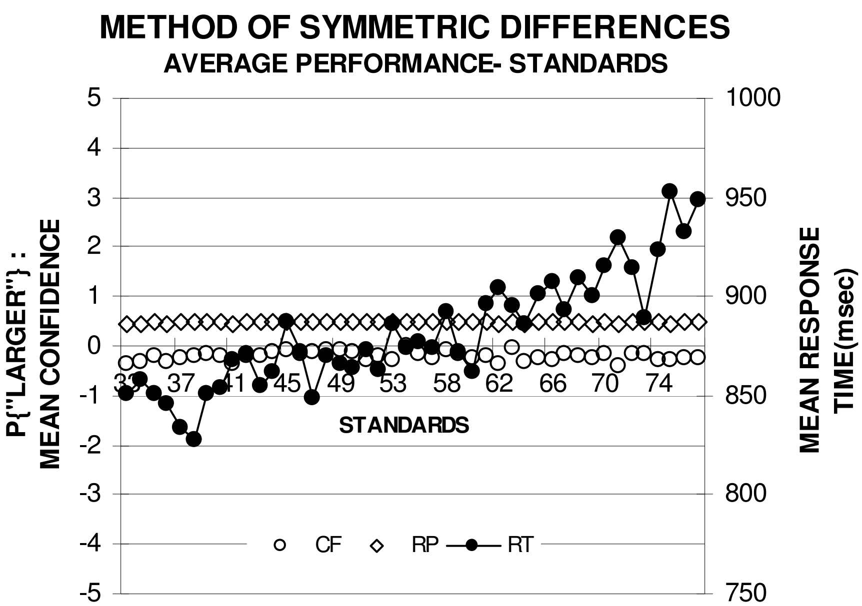


Figure 3. The same results as in Figure 2 but presented as a function of the value of the standard. MSD appears to meet the design criterion of removing response bias based on the standard.

difference. The actual values for proportions and stimulus difference surrounding the point of subjective equality are: (-2ΔS, 0.205), (-1ΔS, 0.300), (1ΔS, 0.486), and (2ΔS, 0.548).

The purpose of MSD was to deny the subject preferential information about the likelihood of a response based on the standard stimulus. To determine whether the design created the desired effect the results were analyzed with respect to the standards against which the variable comparison stimuli were judged. The same performance measures as in Figure 2 appear in Figure 3, with the abscissa scaled according to the value of the standard stimulus. The values of mean confidence (CF) remain flat with an average of -0.21, a very impressive demonstration that the value of the standard does not affect the subjects' response confidence. Second, the average proportion of "LARGER" responses (PF) also remains very flat with an average value of 0.476.

The truly surprising result is the change in mean response time (RT) as a function of the standard. Here the result is not at all flat but fluctuates from low values for the smaller standards (33, 34, …) to quite large values for the largest standards. If anything should convince theoreticians that response proportions do not tell the whole story about response mechanisms, this result should.

The structure of context

Context, it's in the mind of the subject. Words, their meaning, their interpretation, depend on the context in which they are conveyed. Walter Kintsch and Teun van Dijk's analysis of discourse comprehension and Latent Semantic Analysis (Landauer and Dumais, 1994) use context as a source of meaning variation for words. The analysis of words in context gets directly at measures of the ideas in the OED's definitions of context. According information available at the LSA website at the University of Colorado:

> Latent Semantic Analysis is a fully automatic mathematical/statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. It is not a traditional natural language processing or artificial intelligence program; it uses no humanly constructed dictionaries, knowledge bases, semantic networks, grammars, syntactic parsers, or morphologies, etc., and takes as its input only raw text parsed into words defined as unique character strings and separated into meaningful passages or samples such as sentences or paragraphs.
>
> The first step is to represent the text as a matrix in which each row stands for a unique word and each column stands for a text passage or other context. Each cell contains the frequency with which the word of its row appears in the passage denoted by its column. Next, the cell entries are subjected to a preliminary transformation in which each cell frequency is weighted by a function that expresses both the word's importance in the particular passage and the degree to which the word type carries information in the domain of discourse in general.
>
> Next, LSA applies singular value decomposition (SVD) to the matrix. This is a form of factor analysis, or more properly the mathematical generalization of which factor analysis is a special case. In SVD a rectangular matrix is decomposed into the product of three other matrices. One component matrix describes the original row entities as vectors of derived orthogonal factor values, another describes the original column entities in the same way, and the third is a

diagonal matrix containing scaling values such that when the three components are matrix-multiplied, the original matrix is reconstructed. There is a mathematical proof that any matrix can be so decomposed perfectly, using no more factors than the smallest dimension of the original matrix. When fewer than the necessary number of factors are used, the reconstructed matrix is a least-squares best fit. One can reduce the dimensionality of the solution simply by deleting coefficients in the diagonal matrix, ordinarily starting with the smallest. (In practice, for computational reasons, for very large corpora only a limited number of dimensions can be constructed.) From Landauer, T. K., Foltz, P. W., & Laham, D. (1998).

The factor analyst often searches for a basis for personality inventories or intellectual capabilities by correlating the performance on test items across subjects. This is called R-sort and is the basis for understanding performance in terms of an underlying structure for the test items. Often, however, the subjects themselves are of interest. In this case the subjects serve as items and the items as subjects, so that performances are correlated across subjects rather than items. This is Q-sort. This idea was applied by Cartwright, Lee, and Link (1963) to determine the number of factors generated by subjects versus the number of factors derived from an R sort of the items.

To apply this idea to the study of linguistic context requires a transposition of the data used in the LSA method. Transposing the data matrix for LSA exchanges the rows and columns to provide the opportunity to create, through the analysis of words in context, a geometric basis for context. In this way we provide a structure for context, a basis for representing a state of mind, a different approach toward Fechner's goal of inner psychophysical measurement.

References

Cartwright, D. S., Lee, R. L., and Link, S. (1963). Il numero dei fattori e il numero soggetti. Bullettino di Psicologia Applicata, 59-60, 1-9.

Guilford, J. P. (1932). A generalized psychophysical law. Psychological review. 39, 72-86.

Landauer, T. K. and Dumais, S. T. (1994). Latent semantic analysis and the measurement of knowledge. In R. M. Kaplan and J. C. Burstein (Eds), Educational Testing Service Conference On Language Processing Techniques and Technology in Assessment and Education. Princeton. Educational Testing Service.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Link, S. W. (1990a) Modeling imageless thought: the relative judgment theory of numerical comparisons. Journal of mathematical psychology. 34, 2-41.

Link, S. W. (1990b). The psychological design of psychological experiments. In H-G Geissler, M. G. Muller, and W. Prinz (Eds), Psychophysical explorations of mental structures. Hogrefe and Huber.

Thurstone, L. L. (1927). The method of paired comparisons for social values. Journal of abnormal and social psychology. 21, 384-400.

A STATISTICS FOUNDATION COURSE DELIVERED BY COMPUTER

Sandy MacRae
School of Psychology, The University of Birmingham, B15 2TT, England
sandy@statbasics.com

Abstract

Introduction to Statistics provides a full foundation course or a graduate refresher course using sound, voice, animated graphics and a lot of user interaction. It is used in several countries and various disciplines at school and university level. It has a light-hearted approach, aiming to intrigue and entertain as well as educate students worried by statistics. The material covers principles of data-collection, numerical and graphical summaries, correlation and prediction, standard scores, the logic of statistical inference and statistical tests for frequency data and for comparing two groups of scores, whether paired or unpaired. On either a network or a stand-alone computer, it gives each student an individual login that preserves from one session to another records of progress and bookmarks the student creates. It contains many striking animations that can be used in lectures with a projector for memorable communication of concepts. It also allows tutors to construct special-purpose menus to focus on the needs of particular classes or individual students (for project work, say). Tutors can easily monitor student progress, even in large classes, and identify students who are not succeeding. Its web site, www.statbasics.com gives some idea of its scope and style but at the conference you can use the full program and get an evaluation copy.

Writing multimedia content is utterly different from delivering the same material in print. I had written three short textbooks aimed mainly at the large number of students taking A-level examinations in Psychology at the end of secondary education in England. Quantitative methods were prominent in the syllabus but the teachers – let alone the students – often lacked confidence.

From first partial draft to publication took about two years and the typescript was about two inches thick. When my editor offered me the opportunity to convert the material into a multimedia course for computer delivery I accepted eagerly, hugely excited by the possibilities of the medium. Alas, the task proved enormously greater than writing the texts, though the content was very similar. Even with excellent support, it took more than five years and my manuscript was fourteen inches thick. This reflects the fact that almost every screen seen by the user had to be sketched out for the graphic artists with a detailed description of the effects I hoped to achieve.

Partly, the difference is working with editors expert with words compared with working with artists whose skills are mainly visual. But more, the huge flexibility and richness of multimedia creates many more ways to go wrong. For instance, an animation showing how a table of numbers can be transformed into a graph has to be adjusted for speed as well as content. If it is too slow it will seem dreary and if too fast may be incomprehensible. The style adopted has been to begin lengthy animations fairly slowly and speed up once the essential idea of what is happening has been conveyed and every animation can be repeated just by clicking on it if anything remains obscure.

The content is arranged in the eight sections shown in Fig. 1, rather like chapters in a book, but the material can be viewed in any order. Other than going down through the menu hierarchy, buttons on the left provide several other ways to access it. **Study Plans** are created by tutors in the form of a set (up to 20) of individualized menus, each laying out a short syllabus for students to follow. These can assemble sequences from any part of the program to reflect different teaching orders or to direct attention to particular topics. To mark a single screen anywhere in the program, any user can create **Bookmarks**. Typical uses are for students to flag anything they find difficult so they can return later

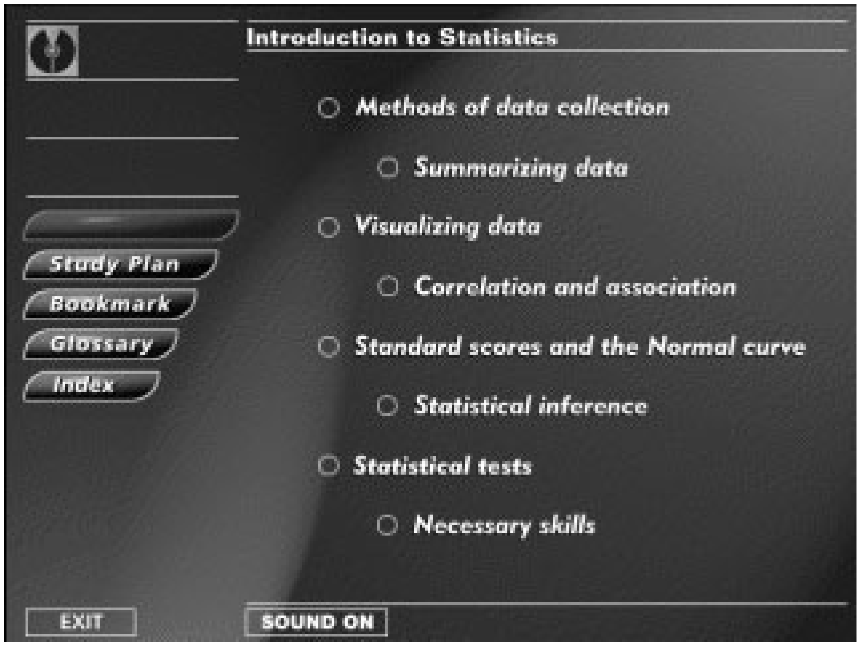


Fig. 1. The main menu. Each selection opens a sub-menu.

to try again or seek advice on it, or for tutors to get instant access to animations they use in lectures. The **Glossary** is a scrollable list of all the program's technical terms, bringing up a brief definition. The **Index** is like the index in a book and allows the user to go straight to the start of a sequence dealing with a particular topic. Together, these ways of reaching the required material or showing students how to reach it combine the flexibility of a free-form tutorial with something like the consistency of a textbook and can be fitted into practically any study program.

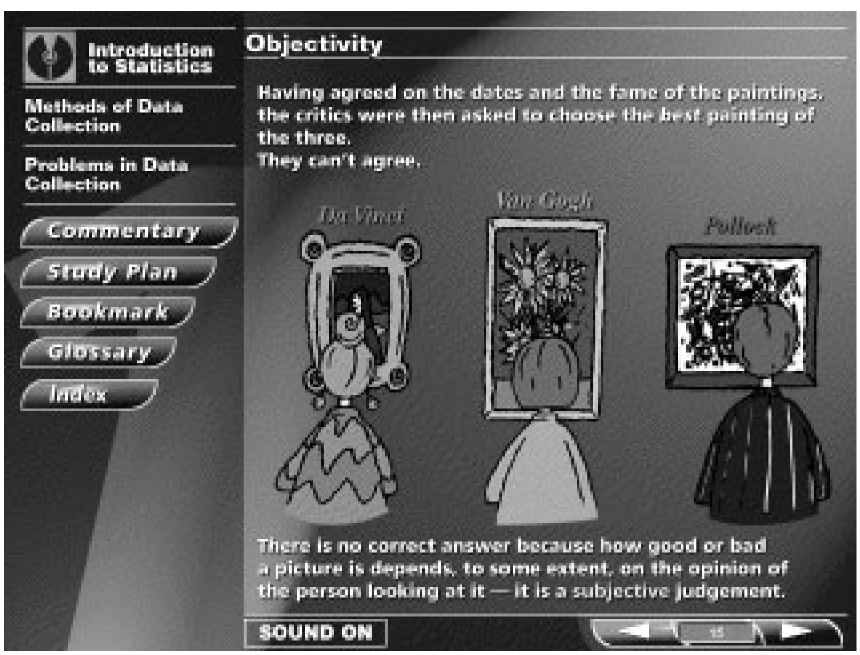


Fig. 2. A teaching screen from Section 1.

The trouble with all this flexibility is that a user might become lost in it, so good navigation guides are needed. Every section has its own dominant colour and the section number (a digit 1 in Figs 2-4) appears unobtrusively in the background. (The colour-coding makes it possible to glance round a large teaching lab and know immediately which sections are being viewed.) The route through the menu hierarchy leading to the current screen is shown at top left and clicking on any of its levels lets the user backtrack to that sub-menu.

Fig. 3. Successive screens introducing nominal scales

A further button, **Commentary**, is available in teaching screens. If the computer has a sound card and phones or speakers, this gives a short, spoken commentary which does not replicate what is shown as text but slightly expands it. The **SOUND ON** button at the bottom of the screen switches sound effects on and off if the equipment is suitable. The program was developed to be entirely comprehensible without sound, though one reviewer mentioned the spoken comments as the only instance he had encountered of ‘added value’ from multimedia presentation.

I aimed from the start to keep the presentation highly interactive. For instance, in the first screen in Fig. 3, the student is invited to choose an answer. If the choice is wrong, a cross briefly appears and fades inviting another try. If correct, a tick appears. Interactions occur frequently in the teaching material, either to engage interest before addressing a topic or at the end of an important explanation to check that the ideas have been assimilated, and take various forms such as drag-and-drop sentence completion or adjustment of a diagram to meet a description.

Fig 4. Successive screens introducing ordinal and interval scales.

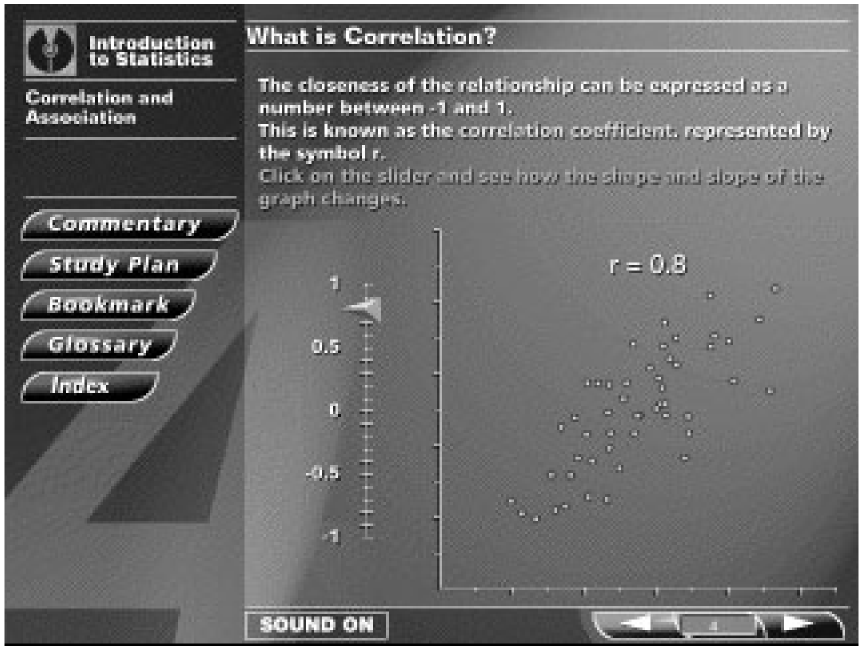


Fig. 5. A scatter plot that can be adjusted by the user while the numerical value of its correlation changes correspondingly between –1 and 1 is often used with a projector for class presentations.

The student is not obliged to respond and can just advance to the next screen by clicking the right-pointing triangle at the bottom, but almost all are intrigued enough to do so. (A user can step forwards and backwards one screen at a time by clicking the triangles in the lower right, or skip several screens by editing the screen number between them.)

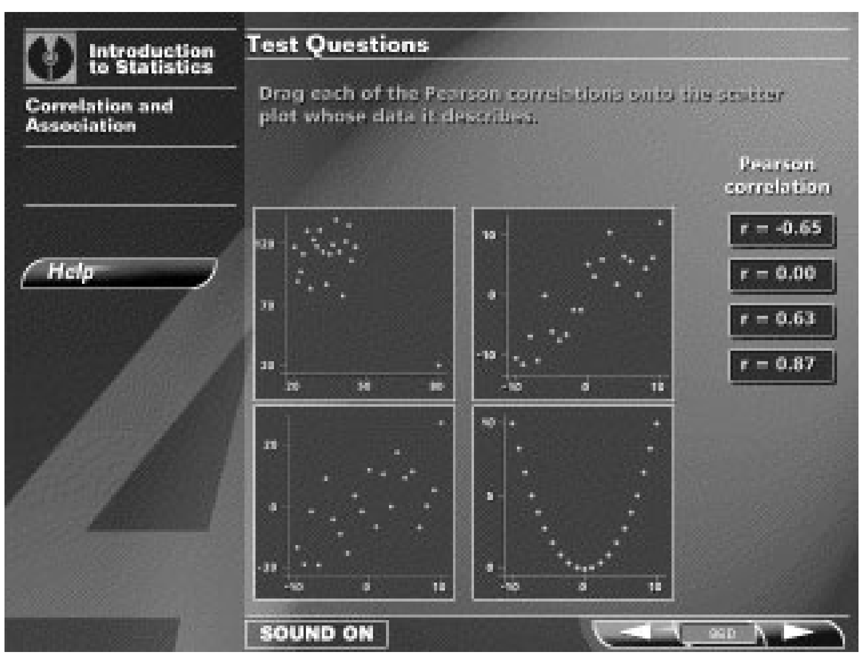


Fig. 6. A test question screen from Section 4 (Correlation and association).

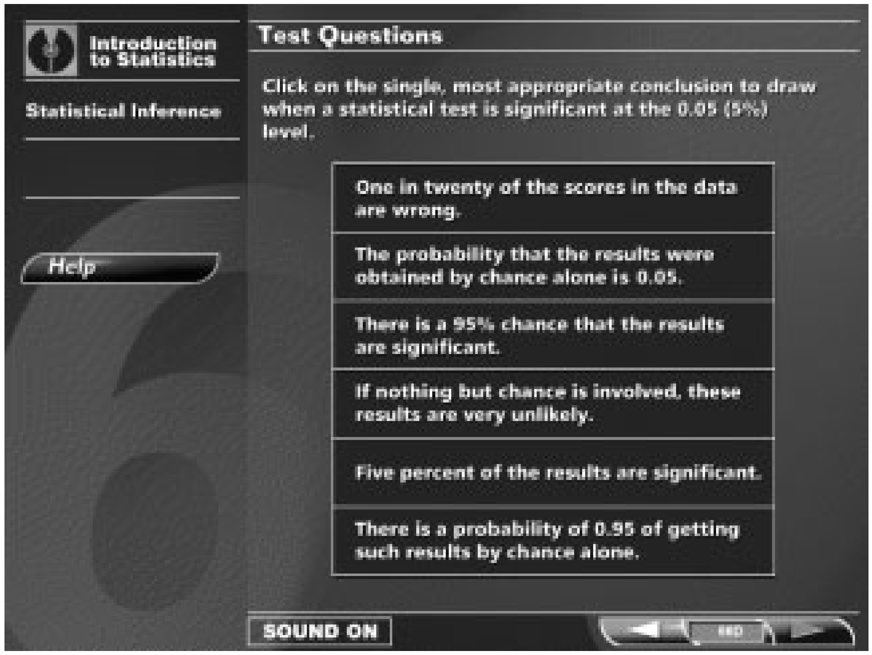


Fig. 7. A test question screen from Section 6 (Statistical inference). Only choice 4 is correct.

At the end of each section there is a set of eight screens containing test questions resembling these tutorial interactions in style but drawing on all the material the student should have mastered. Their level aims to allow a student who has understood the whole section to get all of them right first time. If any errors are made, an opportunity is offered immediately after completing all eight to return and have another go at those that were answered wrongly. The questions are randomized in order and in layout on each presentation so students cannot just learn a spatial pattern of responses. The student can get a hint about how to approach the question by clicking the **Help** button.

Unlike the tutorial interactions, these test scores are recorded and form a record of a student's progress through the course. They are accessible only to the student and to whichever person or group is designated as that student's tutor. All students are expected to succeed and I do not advise using the scores for grading purposes. I know one university course where correct completion of relevant question sections is treated as a course requirement, though in my own courses I only used them to give early warning of students needing special attention.

In the same year that I took early retirement from teaching, I acquired the copyright and after a gruelling, solo development effort produced a revision that works more happily with modern operating systems. Almost all of the original design was retained but some aspects were tidied up and new facilities have been provided. The most important is *Tutor Tools*, a separate program that makes it much easier for tutors to monitor student progress.

Its opening screen is shown in Fig. 8. From the class list (potentially scrolling far below the window) the student Ames has been selected by clicking, so his test results are displayed (8 out of 8 correct answers in Section 3B, 2 out of 8 in Section 5 and nothing else attempted.)

Fig. 8. The main screen of *Tutor Tools.*

The option to **Highlight users** shows different colours for users (tutors or students) who have none, some or all answers correct in whatever Section is selected by radio button. Those currently logged in and those who have never used the program can also be located instantly. There are also facilities to export the results of the whole class to a spreadsheet, save Study Plans or Bookmarks for future use or share them among tutors.

Tutor Tools provides a raft of other facilities for managing the installation, but my focus here is on my particular philosophy and strategies for teaching students in this way. Students often dread the quantitative techniques that are indispensable for any kind of scientific psychology. My aim in class teaching was to arouse interest, give reassurance and entertain, to get them to *want to* understand the ideas I needed to communicate. At first sight, interposing a computer between teacher and student seems to make that more difficult, but my experience is that by capturing my best efforts and the wonderful contributions of several inspired collaborators over a long period and presenting them in concentrated form, it actually improved matters. The only problem was that the program got better student ratings than I did!

Acknowledgement

The indispensable contribution, without which the project could never have been completed, was the inspiration, organizational skill and commitment of Joyce Collins, Books Editor of the British Psychological Society.

INSIGHTS IN SUBJECTIVE SLEEPINESS SCALES

Olivier Mairesse and Peter Theuns
Faculty of Psychology and Educational Sciences, Research Methods and Psychometrics, Vrije Universiteit Brussel, 1050 Brussels, Belgium

Abstract

Self-evaluation category rating scales are widely used instruments in sleepiness research. They are convenient and offer plenty of economic advantages, but unfortunately, they are also associated with psychometric shortcomings. In order to improve the quality of subjective sleepiness assessment, we propose to add a complementory instruction to Borg's CR10 scale (Borg, 1998) to measure perceived effort to stay awake (CR10-SR). Thirty-six male shift-workers performed a 25 min driving simulator test (DriveSim 3.0) between 5.15 and 6.00 AM after a late-night 7 hour shift. Driving performance was assessed using (1) average speed, (2) speed deviation and (3) accident liability. Situational and global sleepiness were assessed with both the Stanford Sleepiness Scale (SSS) and the Epworth Sleepiness Scale (ESS) whereas situational sleep resistance was assessed with the CR10-SR. No significant differences were found in how well SSS or CR10-SR predict average speed, unless the CR10-SR scores were treated as ordinal data. Finally, it seems that CR10-SR predicts speed deviation and accident liability better than the SSS.

Sleepiness is a very strange phenomenon. Everybody knows how *sleepy* feels like, everybody experiences sleepiness daily before sleep onset, and everybody experiences daytime sleepiness from time to time, and yet, as sleep and sleepiness research still are relatively young fields of research, still a lot is to be known.

Different approaches to the quantification of sleepiness have been proposed over the last several decades, hence the amount of assessment tools that have been developed since (for a review, see Cluydts, De Valck, Verstraeten & Theys, 2002). We distinguish three groups of measures for sleepiness: behavioral measures (yawning frequency, eye-blinking,…), direct electrophysiological measures (EEG, EMG,…) and self-evaluation rating scales. For this occasion we will focus on using subjective sleepiness scales. Self-evaluation scales are widely used in sleepiness research and sleep medicine, especially since 1973 when Hoddes, Zarcone, Smythe, Phillips and Dement (1973) presented the Stanford Sleepiness Scale (SSS), a self-report 7-point scale which is used to quantify progressive steps in sleepiness. Numerous studies report that the SSS is a valuable instrument to depict circadian changes in alertness, performance and sleepiness (Johnson, Freeman, Spinweber & Gomez, 1991, Babkof, Caspy & Mikulincer, 1991). Nevertheless, it seems less useful in order to predict individual performance (Herscovitch & Broughton, 1981). Moreover, lay people often misinterpret symptoms of fatigue and tiredness as sleepiness (Cluydts et al. 2002) so that the practicability of self-ratings to assess subjective sleepiness becomes questionable.

The Epworth Sleepiness Scale or ESS (Johns, 1991) is used to measure a *global level* of sleepiness (Johns, 1993). To complete the ESS respondents have to rate, on a 4-point scale, how likely they have been to doze off in eight different daily situations, using a time frame of a few weeks back. The ESS was designed to be robust for situational variations of sleepiness (Miletin & Hanly, 2003) and provides a suitable complementary measure for the assessment of sleepiness. As Cluydts et al. (2002) suggest; sleepiness is not a unitary concept and could be influenced by trait and state-like factors. In this study it is assumed that the SSS measures state-like factors whereas the ESS measures trait-like factors, and that both factors should be implied when investigating levels of sleepiness. In this study we will focus on the assessment of state sleepiness in particular.

In order to improve the quality of situational sleepiness assessment, several issues must be taken into account. Firstly, an operational definition of sleepiness should be the bedrock of the instrument. There

is up till now no consensus about the operationalization of sleepiness. The main issue in this discussion is: “do we consider sleepiness as the *tendency to fall asleep* or rather the *effort* that is *needed to stay awake*?”. Considering the social, health-related and economic relevance of sleepiness research we believe that a practical definition should be endorsed. Most of sleepiness-related traffic accidents happen when the driver fails to stay awake! Excessive daytime sleepiness (EDS) is a major symptom in disorders such as narcolepsy and idiopathic hypersomnia (ICSD-r, American Academy of Sleep Medicine). It implies that high levels of sleepiness interfere with normal daytime functioning, causing the individual to fail withstanding the urge to sleep. Furthermore, we believe that sleepiness is experienced (perceived) most when actually *fighting sleep*.
Secondly, we need a valid instrument, conform to our operationalization of sleepiness. In this case the instrument should measure the perceived the effort to stay awake, as opposed to tiredness and fatigue and offer an clear explanation regarding the symptoms of sleepiness.
Finally, an instrument should be reliable. A single item scale should therefore possess the best possible psychometric properties in order to restrict systematic measurement errors. Borrowing from psychophysical scaling, Borg’s verbally anchored ratio scaling yields category scales that satisfy the properties of a ratio scale. ,The main principle being that in order for a scale to meet the requirements for measurement with ratio properties, a suitable range of intensities should be determined , that are well-placed, and whereby precise verbal anchors should be utilized such that a congruence is obtained between the actual meaning of the verbal anchors and their respective numbers (Borg & Borg, 2002). We believe that Borg’s CR10 could provide a suitable alternative to the classic 7-point single item category scale developed by Hoddes et al.(1973) in measuring situational sleepiness.

Method

The participants in this study are employees of a BASF plant in Belgium, who participated in a broader study aimed to determine which rotation system for night-shift workers would be the most suitable. Thirty-six male shift-workers performed a 25 min driving simulator test (DriveSim 3.0, York Computer Technologies, Kingston, Ontario, Canada) between 5.15 and 6.00 AM after a late-night 7-hour work shift. Before the testing took place, all participants were given the opportunity to take a shower and take a small snack and a non-stimulating drink.

The driving simulator was equipped with an accelerator, brakes and a steering wheel. From the driver’s seat, a motorway road scene with standard road signs and signals was presented on a computer screen. The task consists to maintain a constant speed of 100 km/h and to remain in the right lane a of a straight two-lane route with no bends, stop signs or traffic lights. Opposing traffic and take-over situations were also included. Driving performance was assessed using (1) *average speed*, (2) *speed deviation* (the mean difference, in km/h, between the speed of the vehicle and the posted speed limit, and (3) binary encoded *accident liabilities,* 1 = the car left the road or hit another vehicle at least once during the task, and, 0 else. Prior to the driving task, state and trait sleepiness were assessed using the Stanford Sleepiness Scale (SSS) and the Epworth Sleepiness Scale (ESS). Situational sleep resistance was assessed by using a complementory instruction to Borg’s CR10 scale (Borg, 1998), called here *CR10-SR* (“SR” stands for sleep resistance). The first part of the instructions of the Borg CR-10 scale remain unchanged. The main changes were in order for the second part of the instructions (Fig. 1):

When rating sleepiness, give a number that corresponds to how much effort you have to do at this moment to stay awake. The perception of effort to stay awake is mainly felt as fighting against the urge to fall asleep. You experience sleepiness, yawning, head-nodding and possibly a burning feeling in your eyes. Your eyelids and your head feel heavy and you have trouble remaining alert.

0 ”Nothing at all”. Not sleepy at all means that you are not perceiving any effort to stay awake at all.

1 ”Very weak” means very light. You are experiencing sleepiness very lightly.

3 ”Moderate” is somewhat but not especially hard. You are experiencing sleepiness , but not much. It feels good and it is not difficult to stay awake.

5 ”Strong”. Staying awake is hard, but it is not terribly difficult. The effort to stay awake is about half as intense as “Maximal”.

7 ”Very strong” is quite strenuous. You can still go on, but you have to do a lot of effort to stay awake.

10 ”Extremely strong” – ”Maximal” is an extremely strenuous level. For most people this is the most effort they have put to stay awake.

- *Is ”Absolute maximum – Highest possible”, for example. ”12” or even more.*

Fig 1. Adjusted instructions of the CR10-SR

Results

After deleting outliers based on box plots and constant residuals, 31 subjects were included in the analysis. Participants drove at an average speed of 98,38 ± 2,61 km/h, the mean speed deviation amounted to 6,81 ± 6,39 km/h. Participants reported moderate levels of sleepiness (median SSS score = 3 and mean CR10-SR scores = 3,32 ± 1,19). In order to investigate possible relations between subjective sleepiness measurements and driving ability we first performed bivariate Pearson correlations for *average speed* and *speed deviation* and respective CR10-SR scores. Non-parametric rank-order correlations (Kendall's τ) were performed for *average speed* and *speed deviation* and SSS scores. Results are displayed in Table 1.

Table 1. Correlations between driving ability and situational sleepiness

	SSS	*p* (2-tailed)	CR-10 SR	*p* (2-tailed)
	Kendall's τ		Pearson's *r*	
average speed	- .286	**.046***	- .654	**.000****
speed deviation	.191	.183	.491	**.005****

Conform to Cluydts, De Valck, Verstraeten & Theys (2002), both state and trait sleepiness measures were included in multiple linear regression models in order to predict driving performance:

$$average\ speed_{SS} = \alpha + \beta_1 ESS + \beta_2 D_{1_{sss}} + \beta_3 D_{2_{sss}} + \beta_4 D_{3_{sss}} + \beta_5 D_{4_{sss}}$$

$$average\ speed_{SR} = \alpha + \beta_1 ESS + \beta_2 CR10_{SR}$$

Both models were found to be significant (respectively $F(5,26) = 5{,}761$; $p < .002$ and $F(2,29) = 11{,}228$; $p < .001$). However, in this study ESS scores did not significantly contribute to explain the variance in the models (respectively $B_{1SS} = .103$; $p = .194$ and $B_{1SR} = .072$; $p = .360$) were therefore excluded from the equation.

The variance accounted for by the model with *average speed* as dependent variable and SSS scores as independent categorical variables, corrected for complexity (adj. R^2), equaled .418 ($p < .002$). The same model but with CR10-SR scores as the independent continuous variable showed no important difference: adj. $R^2 = .408$ ($p < .001$). However, our data suggest that participants used the CR10-SR as a category scale, reporting only natural numbers associated with verbal or numerical anchors in distinct categories (Fig. 2).

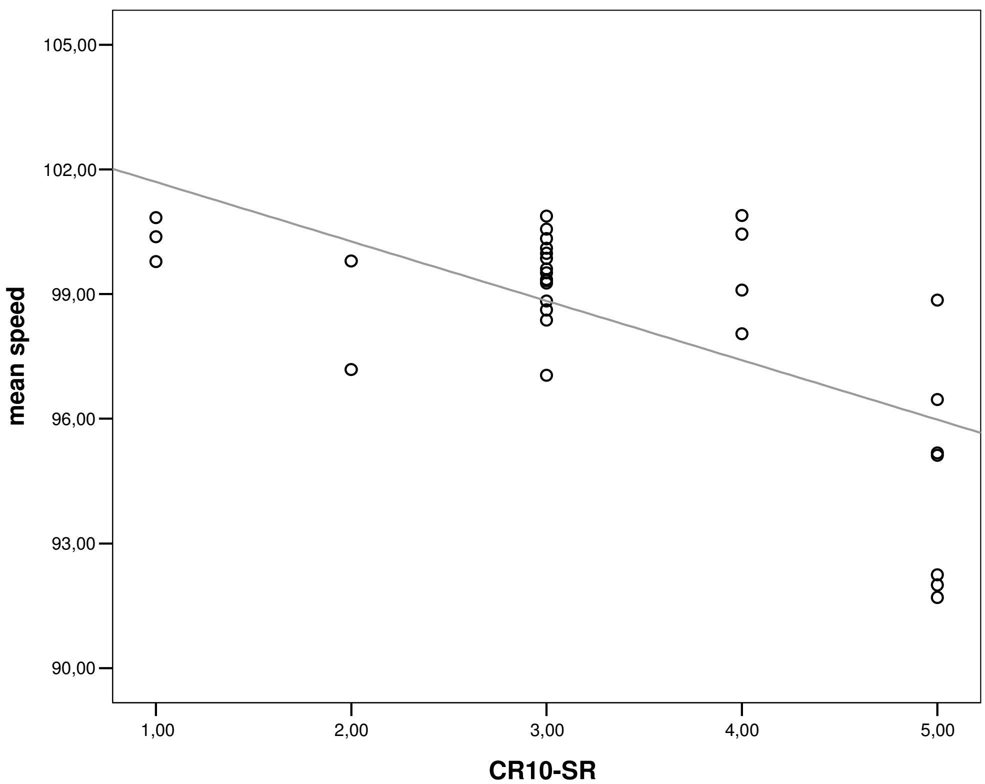


Fig 2. CR10-SR plotted against mean speed. All data is grouped in distinct natural number categories

Considering this, CR10-SR scores were used as independent variable in a linear regression where CR10-SR were recoded into dichotomous variables (dummies). We used the last category (category 5 as the reference category. Our final model explains 63,4% of the variance of *average speed* (R^2 = .634 (p < .001):

$$average\ speed_{SR} = 94{,}51 + 5{,}83D_{1_{CR10SR}} + 3{,}99D_{2_{CR10SR}} + 4{,}94D_{3_{CR10SR}} + 5{,}11D_{4_{CR10SR}}$$

Since no significant correlations were found between s*peed deviation* and SSS scores, we only built a model with *speed deviation* as dependent variable and CR10-SR scores as independent categorical variable. Our model, corrected for complexity, is significant (F= 4,289; p< .01) and explains 30,5% of the variance of the dependent variable (adj. R^2= .305). One of our parameters was not significant (category 2: $B_{2_{CR10SR}}$ = -7,00; p = .113) and therefore excluded from the equation. Our final model is significant (F= 4,54; p< .02) but explains only 26,1% of the variance (adj. R^2= .261):

$$speed\ deviation_{SR} = 12{,}47 - 8{,}5D_{1_{CR10SR}} - 8{,}14D_{3_{CR10SR}} - 6{,}92D_{4_{CR10SR}}$$

In order to predict *accident liability* logistic regressions were performed. CR10-SR scores as independent categorical variables significantly predict deviations from the road or hits with other vehicles (χ^2 = 9,716; p < .05), where SSS scores failed to be significant predictors.

Discussion

Wilson once stated that *"the ordinal level of measurement prohibits all but the weakest inferences concerning the fit between data and a theoretical model formulated in terms of interval variables... The task of developing valid, reliable interval measurement is not a technical detail that can be postponed indefinitely…"* (1971, see Roberts 1994). That was about 35 years ago and still today category scaling is probably the most frequently used format of rating scales (Rohrmann, 2002). Despite the fact that ordinal scales, due to their limited measurement level, are suspected to yield data of inferior quality (Hofmans et al. 2005), our data suggest that replacing such ordinal scales with more complex formats is not as obvious to every person as we would like it to be. Respondents often prefer category rating scales because they experience those as convenient and supportive to express their true opinion (Garland, 1990). Visual analogue scales (VAS), for instance, have also been used in sleep research. A well-constructed VAS provides data of ratio measurement level but requires some understanding of geometry and the ability to map an abstract phenomenon onto a scale. Even the concept of such a simple scale can be foreign to rural people or to the poorly educated (Maldonado, Bentley & Mitchell, 2004). The issue of trying to optimize measurement level in scaling comes with a rather important practical covariates; the familiarity and ease of use of the scale.

A suitable explanation for the rather disappointing differences found between SSS an CR10-SR scores in predicting *mean speed* could thus be found in the particular composition of our sample. All tested subjects had a rather low level of education and may have found it difficult to use Borg's scale properly. Another possible explanation is that the second part of the instructions (see Fig. 1) instigates subjects to give numerical responses corresponding to the explicative verbal descriptors. In our sample, it is indeed plausible that the participants used the second part of the instruction to make their judgements instead of using the actual CR10 scale displayed with numbers and verbal indicators.

Still, in a practical sense, when CR10-SR data are treated as if they were ordinal measures, the instrument still seems to predict the dependent variables *average speed*, *speed deviation* and *accident liability* more accurately than the SSS. Those differences cannot be attributed only to the difference in available categories –seven for the SSS, ten (?) for the CR10-SR since only five categories were actually used in both scales. These results suggest that situational sleep resistance measured by the CR10-SR could predict driver performance more accurately than the classical measure of subjective situational sleepiness. In order to optimize the assessment of situational subjective sleepiness further meta-analyses of the CR10-SR are needed, as well as validation studies using within-subjects or cross-sectional designs.

Bibliography

Babkoff H., Caspy T., Mikulincer M. (1991) Subjective sleepiness ratings: the effects of sleep deprivation, circadian rhythmicity and cognitive performance. *Sleep,* 14(6):534-9.

Borg G. (1998) Borg's perceived exertion and pain scales. *Human Kinetics*, Leeds, United Kingdom.

Borg E., Borg G. (2002) A comparison of AME and CR100 for scaling perceived exertion. *Acta Psychologica* ,109(2):157-75.

Cluydts R., De Valck E., Verstraeten E., Theys P. (2002) Daytime sleepiness and its evaluation. *Sleep Medicine Reviews*, 6: (2) 83-96.

Garland, R. (1990). A Comparison of Three Forms of the Semantic Differential Marketing Bulletin, 1, 19-24. Retrieved June 29, 2004, from http://marketing-bulletin.massey.ac.nz/article1/V1Article4.pdf

Herscovitch J., Broughton R. (1981) Sensitivity of the Stanford sleepiness scale to the effects of cumulative partial sleep deprivation and recovery oversleeping. *Sleep*,4(1):83-91.

Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC. (1973) Quantification of sleepiness: a new approach. *Psychophysiology,*10(4):431-6.

Hofmans J., Cools W., Baekelandt S., Theuns P. (2005) Bias and changes in perceived intensity of verbal qualifiers effected by scale orientation. *Journal of Market Research* (accepted).

Johns, M.W. (1993) Daytime sleepiness, snoring, and obstructive sleep apnea. The Epworth Sleepiness Scale *Chest.* 103(1):30-6.

Johns M.W. (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep*, 14(6):540-5.

Johnson L.C., Freeman C.R., Spinweber C.L., Gomez S.A. (1991) Subjective and objective measures of sleepiness: effect of benzodiazepine and caffeine on their relationship. *Psychophysiology,* 28(1):65-71.

Maldonado C.C., Bentley A.J., Mitchell D. (2004) A pictorial sleepiness scale based on cartoon faces. *Sleep*, 27(3):541-8.

Miletin M.S., Hanly P.J. (2003) Measurement properties of the Epworth sleepiness scale. *Sleep Medicine*, 4(3):195-9.

Roberts, F. S. (1994) Limitations on conclusions using scales of measurement. In A. Barnett, S. Pollock & M. Rothkopf (Eds.), *Operations research and the public sector* (pp. 621-671). Amsterdam: Elsevier.

Rohrmann, B. (2002) Verbal qualifiers for rating scales: Sociolinguistic considerations and psychometric data. Project for University of Melbourne.

DEFINING THE DIMENSION OF A PSYCHOLOGICAL VARIABLE USING DIMENSIONAL ANALYSIS

S. Alex Marinov
Department of Health, MMF Inc, 300-150 Henry Avenue, Winnipeg, MB, R3B 0J7, Canada, marinov@escape.ca

Abstract

In the first part of this study we introduce two fundamental concepts of physics - *physical dimension* and *the principle of dimensional homogeneity*, and describe the method of *dimensional analysis* - one of the most powerful tools of physics. Then, we ask the central question of this article: Why such powerful concepts and methods have not been used beyond physics, including in psychophysics and psychology? The answer given is: Because of the lack of notion of dimension of a non-physical quantity. In the second part we introduce a method of assigning dimensions to non-physical variables, and argue that such dimensions, regarded as *phenomenological dimensions*, allow for meaningful interpretation. The method, referred to as *Reversed Dimensional Analysis* (RDA), is based on the principal of dimensional homogeneity and the technique of dimensional analysis. Finally, we provide two examples of applying RDA to real data obtained in psychophysical and short-term memory experiments.

The concept of dimension of a quantity can be traced back to the ideas of geometrical dimensions developed during the Greek antiquity (Martins, 1981). In their system of geometry lines had one dimension, surfaces two, and solids three. In the beginning of the 19-th century Fourier (1822/1955) explicitly introduced the notion of physical dimension as an analogue and extension of the mathematical concept of dimension. Fourier also introduced the division of physical dimensions into two categories - primary (basic or fundamental) and secondary (derived) dimensions - the latter being defined either by a definition of the corresponding quantity or by a physical law which provides a relation between the quantity and other quantities of primary dimensions. For example, if, as commonly accepted in mechanics, the dimensions of mass (**M**), length (**L**) and time (**T**) are used as a set of primary dimensions, then the dimension of density is defined as ML^{-3} according to the definition of density (the ratio of mass to volume), while the dimension of force is defined as MLT^{-2} according to the second law of Newton (**F=ma**). Generally, any secondary dimension in mechanics can be expressed as a monomial of the kind $\mathbf{M^{\mu}L^{\lambda}T^{\tau}}$, where the exponents μ,λ,τ are small integers: $\mathbf{-1\leq\mu\leq+1}$, $\mathbf{-4\leq\lambda\leq+4}$ and $\mathbf{-3\leq\tau\leq+2}$ (Krantz et al., 1971). Such a monomial is referred to as a dimensional formula of the corresponding quantity. The most widely used measurement system - *Système International* (SI) - consists of 7 quantities of primary dimensions: mass, length, time, electrical current, thermodynamic temperature, amount of substance and luminous intensity. The dimension of each of these quantities is independent of the dimensions of the rest in the sense that it cannot be expressed as a product of powers of the remaining primary dimensions. On the contrary, the dimensions of all currently known quantities of secondary dimensions can be expressed as such products. Physical dimensions are viewed as: "***names** of particular classes of similar scales for the measurement of quantities*" (Ellis, 1966, p. 139), "***physical attributes** that have ratio scale measures*" (Krantz et al., 1971, p.455), "[being related to] ***the nature of the physical quantity without reference to its magnitude***." (Pankhurst, 1964, p.20), or *"**qualitative aspects** ... that serve to identify the nature or type of the characteristics of the objects"* (Munson, Young & Okiishi, 1990, p.5). The common denominator of such definitions is the view that physical dimensions are qualitative denotations (names) of different properties of physical objects or phenomena.

The genesis and metamorphoses of the notion of dimension were historically paralleled by the genesis and metamorphoses of another fundamental idea of physics - the notion of dimensional homogeneity. The latter came into being in considering what can and what cannot be compared quantitatively. *Homogenea homogeneis comparare* ("Homogeneous quantities must be compared with homogeneous quantities") postulated François Viète in 1591 in formulating his first law of geometric equalities, which stated that only geometric quantities of same dimensions can enter a geometric equation (Fregulia, 2001). Consequently, he was only interested in solving "geometrically possible" equations such as $\mathbf{X^3+A^2X=A^2B}$, where all three quantities have a dimension of length (**L**) and, respectively, all three terms have a dimension of volume ($\mathbf{L^3}$). On the contrary, he was not interested in "geometrically impossible" equations such as $\mathbf{X^3+X=1}$, as adding volume to length was considered by him as meaningless. More than two centuries later Fourier (1822/1955) extended the principle of dimensional homogeneity to physical quantities and regarded it as "*the equivalent of the fundamental lemmas* [of geometry] *which the Greeks have left us without proof*". According to him "*every ... magnitude or constant has one dimension proper to itself, and that the terms of one and the same equation could not be compared, if they had not the same exponent of dimensions*". For example, the equation describing a falling body, $\mathbf{s=(1/2)gt^2}$, is dimensionally homogeneous as the two terms of it have the same dimension: $\mathbf{[s]=L}$ and $\mathbf{[gt^2]=(L/T^2)T^2=L}$, where brackets (**[ ]**) are used to denote the dimension of the quantity, and **L** and **T** have to usual meaning of dimensions of length and time. Attempts have been made to justify the principle of dimensional homogeneity, in essence an axiom, by resorting to rather metaphysical concepts such as *meaningfulness*. About 30 years ago, Luce (1978) published a theorem which stated that "*dimensionally invariant numerical laws correspond to meaningful quantitative relations*". Most recently Falmagne (2004) used the principle of dimensional homogeneity as a basis for introducing a fundamental principle (an axiom) of scientific laws referred to as an invariance principle "called *meaningfulness*" and rooted in "*the common practice requiring that the form of a scientific law must not be altered by a change of the units of the measurement scales*", the latter statement being an alternative form of wording the principle of dimensional homogeneity.

Dimensional analysis (DA) - a method based on the principle of dimensional homogeneity - is justly regarded as the most fundamental and most universal method of physics (Smith, 1993). DA serves a number of purposes: checking physical equations for errors; reducing, without lost of information, the number of dimensional variables describing a phenomenon to a smaller number of dimensionless variables; and modeling physical phenomena, among others (Barenblatt, 1987; Brigman, 1922). A major tool in DA is the so-called *Π-theorem* (Buckingham, 1914). It states that <u>if</u> $\mathbf{f(X_1, X_2 ... X_n)=0}$ is the relation between **n** dimensional variables and constants, where $\mathbf{X_1, X_2, \dots X_n}$ is a complete set of variables/constants relevant to the phenomenon being investigated, <u>then</u> such a relation can be reduced to a relation between **n-k** dimensionless variables $\mathbf{F(\Pi_1, \Pi_2 ... \Pi_{n-k})=0}$, where **k** is the number of the dimensional variables/constants which have independent dimensions (i.e., their dimensions cannot be expressed by the dimensions of the remaining variables), and the Π-terms are dimensionless products of powers of the initial dimensional variables and constants: $\mathbf{\Pi=X_1^a X_2^b \dots X_n^h}$, where the exponents **a, b … h** are appropriately chosen small integers. (The indexes for the particular $\mathbf{\Pi_i}$'s are omitted.) In many practical cases the exponents are chosen by inspection, although general methods have also been developed (Boyle, 1986; Sabersky, Acosta & Hauptmann, 1989). The typical way of applying DA to a problem is, in the first place, creating a complete list of dimensional variables and constants relevant to the phenomenon being investigated (DA offers no help at this stage), secondly, applying the Π-theorem (at this stage, as it will be illustrated below, DA may provide some clues regarding omitted relevant variables/constants and/or included irrelevant ones), and, thirdly, determining the explicit form of **F** via fitting data. (DA *per se* does not allow to find the explicit form of this function except for the simplest but not so rare case of **n-k=1** which results in $\mathbf{\Pi_1=Const}$).

Solving the problem of simple pendulum is commonly used as an illustration of DA (cf. Barenblatt, 1987; Brigman, 1922; Krantz et al., 1971). Intuitively, the list of quantities relevant to the swing of a pendulum shall include: the time (**t**) and the angular amplitude (**θ**) of swing, the length (**l**) and the mass (**m**) of the pendulum, and the acceleration of gravity (**g**). The dimensions of these quantities are respectively: $[\mathbf{t}]=\mathbf{T}$, $[\boldsymbol{\theta}]=\mathbf{1}$ (dimensionless), $[\mathbf{l}]=\mathbf{L}$, $[\mathbf{m}]=\mathbf{M}$ and $[\mathbf{g}]=\mathbf{LT^{-2}}$. The relation between these five quantities can be written as $\mathbf{f(t,\theta,l,m,g)=0}$. As there are three quantities of independent dimensions (**t**, **l** and **m**), applying the Π-theorem results in $\mathbf{F(\Pi_1, \Pi_2)=0}$, where $\mathbf{\Pi_1=\theta}$ and $\mathbf{\Pi_2=t^2l^{-1}g}$. A simple inspection shows that both Π-terms are dimensionless. This example illustrates two of the purposes of DA. First, a quantity, the mass (**m**), intuitively considered to be relevant to this problem, was in fact eliminated as such because it could not combine in any manner with the remaining quantities in order to enter a Π-term. Secondly, the initial problem of five dimensional variables/constants was reduced to a problem of only two dimensionless variables. If, for instance, we are now interested in the relation between the time of swing of the pendulum (**t**) and the other relevant variables, then $\mathbf{\Pi_1=\theta}$ and $\mathbf{\Pi_2=t^2l^{-1}g}$ can be substituted in $\mathbf{F(\Pi_1, \Pi_2)=0}$ and the latter be rewritten as $\mathbf{t^2l^{-1}g=F(\theta)}$, or in more conventional form as $\mathbf{t=(g/l)^{-1/2}F(\theta)}$, where **F** stands for any function. Analytical solutions and experimental data confirm this result (Brigman, 1922). The final step in solving this problem with the help of DA would be fitting measurement data and determining the explicit form of **F**.

Having introduced the concept of dimension, the principle of dimensional homogeneity and the method of dimensional analysis, we now ask the central question of this article: Why such powerful concepts and methods have not been used beyond physics, including in fields as psychophysics and psychology? The answer is rather simple: Because of the lack of notion of dimension of a non-physical quantity.

In the second part of this article we introduce a method of assigning dimensions to non-physical variables, and argue that such dimensions, regarded as *phenomenological dimensions*, allow for meaningful interpretation. The method, referred to as Reversed Dimensional Analysis (Marinov, 2004) is based on the principal of dimensional homogeneity and the technique of dimensional analysis, however, unlike the traditional use of DA, it essentially utilizes data.

The following is a brief discourse of the logic of Reversed Dimensional Analysis (RDA): We first note, that the principle of dimensional homogeneity makes sense only when applied to quantities of known dimensions. Respectively, dimensional analysis can only be applied to such quantities. If this is the case, as shown above, the relation between the relevant dimensional variables $\mathbf{f(X_1, X_2 ... X_n)=0}$ can be transformed into a relation between dimensionless variables $\mathbf{F(\Pi_1, \Pi_2 ... \Pi_{n-k})=0}$, where the Π-terms have the form $\mathbf{\Pi=X_1^a\, X_2^b \ldots X_n^h}$, and all exponents **a, b … h** are known small integers. Up to here no data are required. Data become important only when we want to find the explicit form of **F**, which is practically achieved by fitting such data with various functions and choosing the one that provides the best fit. If, however, even one of the quantities $\mathbf{X_1, X_2, \ldots X_n}$ has unknown dimension, then DA cannot be applied in order to transform **f** into **F**, in the sense that the form of at least one of the Π-terms could not be determined. Provided $\mathbf{X_1, X_2, \ldots X_n}$ is a complete set of relevant variables, the only inference that can be made at this stage (before involving data) is that the unknown dimension must depend on the known dimensions, and therefore the number of Π-terms, **n-k**, will not be affected. (The latter statement is not necessarily true in case of more than one variable of unknown dimension, but for the sake of simplicity, we limit this consideration to case of only one such variable.) If, for example, $\mathbf{X_1}$ is the variable of unknown dimension, then its dimension will be linked to the dimensions of the other variables by the expression $\mathbf{[X_1]=[X_2]^{b^*}[X_3]^{c^*}\ldots[X_n]^{h^*}}$, where, however, **b*, c* ... h*** are unknown exponents, as opposed to the case when all variables have know dimensions. (Asterisks are used to distinguish these exponents from those that appear in the above description of DA.). If we are now interested in finding the dimension of $\mathbf{X_1}$, then it is obvious that the only way to do so is to substitute the expression for $\mathbf{[X_1]}$ into $\mathbf{F(\Pi_1, \Pi_2 ... \Pi_{n-k})=0}$ and try to fit data with various functions, considering the exponents **b*, c* ... h*** as parameters of fitting along with the ordinary fitting parameters (such as **A** and **B** in fitting data with a straight line **Ax+B**). Two important aspects of the described approach have to be emphasized. Firstly, while in the case of DA the function **F** always exists and the purpose of fitting data is only to find its explicit form, in the case of RDA it may or may not exist. Thus in the latter case the purpose of fitting data is also to prove the existence of **F**, of course, in statistical sense. Secondly, if we do not limit the values of **b*, c* ... h*** to integers only, then we introduce a notion of a non-integer dimension which is fundamentally different from the notion of dimension in physics. (Recall, that all exponents in the dimensional formula of a quantity are small integers.) Fortunately, the late 20 century' studies in *fractal geometry* (Mandelbrot, 1983), have shown not only that geometric quantities of non-integer dimensions (viz. *fractals*) allow for meaningful interpretation, but that in fact this geometry provides a more adequate description of a large variety of shapes found in nature. Moreover, Marinov (2001, 2006) showed that a fractal dimension is a special case of a dimension obtained by RDA. To distinguish such dimensions from other kinds of dimensions, and to emphasize the role of data in obtaining dimensions by RDA, the term *phenomenological dimension* was suggested (Marinov, 1999).

In the last part of this article we provide two examples of applying the method of RDA to real data obtained, respectively, in psychophysics and short-term memory studies. Readers interested in more detailed discussion of these examples may consult Marinov (2004).

Consider Stevens' psychophysical law for time duration, $\mathbf{\Psi=C\Phi^{m}}$, where $\mathbf{\Psi}$ is the magnitude of sensation, $\mathbf{\Phi}$ is the stimulus magnitude, and **C** and $\mathbf{m\approx1.15}$ are parameters usually obtained by fitting experimental data (Stevens, 1959). Within the framework of RDA, if $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are considered as the only variables relevant to the phenomenon being investigated and there is no reason to introduce dimensional constants, then the dimension of one of these variables must depend on the dimension of the other variable, and therefore the dimensional relation between these two variables, $\mathbf{\Psi=f(\Phi)}$, can be reduced to only one dimensionless variable, $\mathbf{\Pi=\Psi\Phi^{m^*}}$, where, according to the Π-theorem, $\mathbf{\Pi=Const}$. If, for instance, the dimension of the magnitude of sensation ($\mathbf{\Psi}$) is considered unknown while the dimension of the stimulus magnitude ($\mathbf{\Phi}$) is assumed to be known, then, as far as $\mathbf{\Pi}$ is dimensionless, the dimension of $\mathbf{\Psi}$ can be expressed by the dimension of $\mathbf{\Phi}$ according to the expression $\mathbf{[\Psi]=[\Phi]^{-m^*}}$, where $\mathbf{-m^*}$ is an unknown exponent. (The minus sign plays no special role but is rather kept for consistency.) As previously argued, the only way to find this exponent is to resort to data. In the case of $\mathbf{\Psi}$ and $\mathbf{\Phi}$ the fitting procedure will be exactly the same as the one used in fitting psychophysical data in order to estimate **C** and **m** in $\mathbf{\Psi=C\Phi^{m}}$. Thus, using the commonly adopted value for time duration, $\mathbf{m\approx1.15}$, the dimension assigned to the magnitude of sensation will be $\mathbf{[\Psi]=[\Phi]^{-m^*}=T^{1.15}}$, where **T** is the dimension of time. (Following the above denotations: $\mathbf{\Pi=\Psi\Phi^{m^*}=Const}$, hence $\mathbf{\Psi=Const.\Phi^{-m^*}}$, and therefore $\mathbf{-m^*}$ in the latter relation corresponds to **m** in $\mathbf{\Psi=C\Phi^{m}}$).

It would be reasonably to ask: What is the advantage of assigning a non-integer dimension of time, $\mathbf{T^{1.15}}$, to the magnitude of sensation ($\mathbf{\Psi}$), as opposed to other possible alternatives such as: (i) considering $\mathbf{\Psi}$ as having the dimension of $\mathbf{\Phi}$, i.e. **T**, and respectively attributing a dimension of $\mathbf{T^{-0.15}}$ to the parameter **C**; <u>or</u> (ii) considering $\mathbf{\Psi}$ as dimensionless, and respectively attributing a dimension of $\mathbf{T^{-1.15}}$ to **C**; <u>or</u> (iii) simply ignoring the issue of dimensions of $\mathbf{\Psi}$ and **C**? The disadvantage of (i) and (ii) is apparently the fact that a dimensional constant of no clear meaning is introduced. As stressed by Brigman (1922), such constants are to be strongly avoided. The disadvantage of (iii) is that two quantities of different dimensions (viz. qualities) are equated, that is a clear violation of the principle of dimensional homogeneity and therefore meaningless in light of the concept of *meaningfulness* of scientific laws advanced by Luce (1978) and Falmagne (2004). Obviously, RDA overcomes both problems. The problem left, of course, is attributing a meaning to the magnitude of sensation as a quantity of non-integer dimension. At this time we can advance at least two arguments in favor of this approach. In the first place, no fundamental restrictions on existence of non-integer dimensions have ever been imposed (cf. Krantz et al., 1971). On the contrary, as Barenblatt (1987) and many others (cf. Mandelbrot, 1983) show, in the case of complex geometric patterns the only adequate description of such patterns is attributing to them a non-integer dimension of length, $\mathbf{L^{k^*}}$, where **k*** may have any value between 0 and 3, including non-integer ones. Then, why such a complex phenomenon as perception of physical magnitudes cannot be characterized by quantities of non-integer dimensions? The second argument in favor of assigning a dimension to $\mathbf{\Psi}$ is that, if such a dimension reflects any essential property (Ellis, 2001) of perception of time duration, then we may, or at least it is attractive to try, to trace this dimension in other psychological phenomena involving perception or processing of time intervals. The next paragraph provides an encouraging example.

In a Peterson & Peterson' (1959) type of experiment Hellyer (1962) studied the Proportion of items (consonant trigrams) correctly recalled (**P**) as a function of two variables - Recall delay interval (**R**) and Number of presentations (**N**). The relation between these three variables can be written as **P=f(R,N)** where, according to Hellyer, the dimensions of the variables are **[P]=1** (dimensionless), **[R]=T** (dimension of time) and **[N]=1** (dimensionless). Apparently, this relation is dimensionally inhomogeneous unless at least one dimensional constant, say $\mathbf{C_H}$, with a dimension of the kind $\mathbf{[C_H]=T^{t^*}}$, is introduced, where the exponent **t*** will depend on the explicit form of **P=f(R,N)**. Within the frame work of RDA, if we now consider **N** as a quantity of unknown dimension which in this particular case may only depend on the dimension of **R**, then applying the Π-theorem to **P=f(R,N)** results in $\mathbf{\Pi_1=F(\Pi_2)}$, where $\mathbf{\Pi_1=P}$, $\mathbf{\Pi_2=R^{r^*}N}$, and **r*** is an unknown exponent that can only be found by resorting to experimental data. As $\mathbf{\Pi_2}$ is a dimensionless variable, the dimension of **N** can be expressed by the dimension of **R** according to the expression $\mathbf{[N]=[R]^{-r^*}}$. Avoiding details, which can be found in Marinov (2004), the fitting of Hellyer's data with various functions resulted in **r*=-1.13**. Respectively, the dimension of **N** is $\mathbf{[N]=[R]^{-r^*}=T^{-(-1.13)}=T^{1.13}}$. Thus, the phenomenological dimension of the variable Number of presentations in Hellyer's data is practically the same as the phenomenological dimension of the magnitude of sensation in Stevens' psychophysical law for time duration ($\mathbf{T^{1.15}}$). In other words, in two different psychological experiments - magnitude estimation and memory recall - it was found that variables related to processing of time exhibit the same phenomenological dimensions. It can, of course, be argued that **N** (Number of presentations) is an external independent variable and therefore has nothing to do with the way time intervals are processed by the subject. With respect to the variable *per se*, this is correct. The computation of the phenomenological dimension of this variable, however, involves all three variables, including **P** (Proportion of items correctly recalled) which is in fact the measure of the response of the subject. As found in Hellyer's data, this response is affected by both external variables **R** and **N**. In other words, the time intervals involved in the Hellyer's experiment, essentially affect the way the subject recalls and presumably memorizes items.

Based on the two examples considered, it can be claimed that the phenomenological dimension related to processing time intervals is the same in two different psychological experiments. Is this, however, always the case so that it can be claimed that the dimension of the quantity attributed to processing time has been discovered, and that this dimension is $\mathbf{T^{1.15}}$? (More accurately, due to the statistical nature of obtaining the number for the exponent, it should be said "about $\mathbf{T^{1.15}}$", where "about" applies to the numerical value of the exponent). The answer is rather no (Marinov, 2004). In fact, it would be surprising, was it otherwise, in light of the high variability of data obtained in psychological experiments. Yet, there are indications that RDA produces rather similar results when applied to data obtained in similar conditions (Marinov, 2004), and therefore there is a hope that a phenomenological dimension may capture the essence of the property being measured.

References

Barenblatt G.I. (1987) *Dimensional analysis*. New York: Gordon & Breach

Boyle W.P. (1986) *Applied fluid mechanics*., Toronto etc.: McGrow - Hill

Bridgman P. (1922) *Dimensional analysis*. New Haven: Yale University Press

Buckingham E. (1914) On physically similar systems: Illustrations of the use of dimensional equations. *Physical Review*, 4, 345-376

Ellis B. (1966) *Basic concepts of measurement*. Cambridge: Cambridge University Press

Ellis B. (2001) *Scientific Essentialism*. Cambridge: Cambridge University Press

Falmagne J.C. (2004) Meaningfulness and Order-Invariance: Two Fundamental Principles for Scientific Laws, *Foundations of Physics*, 34 (9), 1341 - 1384

Fourier J. (1822/1955), *Theorie Analytique de la chaleur*. Gauthier-Villars, [English translation: Fourier J. (1955) *The Analytical Theory of Heat*. Dover, 1955]

Freguglia P. (2001) About the law of dimensional homogeneity between the 16th and the 17th century. Paper presented on the *International Workshop: Science and Power during the Cold War in the European Periphery*, Barcelona, November 1-3, 2001

Hellyer S. (1962) Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental Psychology*, 64, 650

Krantz D.H., Luce R.D., Suppes P., & Tversky A. (1971) *Foundations of measurement: Vol. I. Additive and polynomial representations*. New York: Academic Press

Luce R.D. (1978) Dimensionally invariant numerical laws correspond to meaningful quantitative relations. *Philosophy of Science*, 45, 1-16

Marinov S.A. (1999) Dimensions as Quantities, In B. Kreplin, S. Rudolph & S. Bruckner (Eds.) *Proceedings of the 2-nd International Workshop on Similarity Methods*, November 25-26, 1999, Stuttgart, Germany, 111-128

Marinov S.A. (2001) Fractal dimensions as a special case of phenomenological dimension., In B. Kreplin, S. Rudolph & S. Bruckner (Eds.) *Proceedings of the 4-th International Workshop on Similarity Methods*, November 05-06, 2001, Stuttgart, Germany, 117-127

Marinov S.A. (2004) Reversed Dimensional Analysis in Psychophysics. *Perception & Psychophysics*, 66(1), 23-37

Marinov S.A. (2006) A Generalized Fractal Dimension Based on Dimensional Analysis. Presentation, *Fractal 2006 - Biannual Conference*, February 12-15, 2006, Vienna, Austria [accepted]

Martins R.A. (1981) The origin of dimensional analysis. *Journal of the Franklin Institute,* 311, 331-337

Munson B.R., Young D.F. & Okiishi T.H. (1990) *Fundamentals of fluid mechanics*. New York: Wiley

Pankhurst R.C. (1964) *Dimensional analysis and scale factors*. New York: Reinhold

Peterson, L.R., & Peterson, M.J. (1959). Short-term memory retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198

Sabersky R.H., Acosta A.G., Hauptmann E.G. (1989) *Fluid flow: A first course in fluid mechanics*., New York: Macmillan

Smith E. (1993) Fundamental Concepts of Century Physics, *University of Texas*, http://www.ph.utexas.edu/~gleeson/httb/httb.html [accessed: July 2005]

MENTAL ROTATION OF HANDS:
COMPARING MOTOR ABLE AND CEREBRAL PALSY SUBJECTS

Isabel Catarina Martins[1], Armando Mónica de Oliveira[1], Michel-Ange Amorim[2]
[1]University of Coimbra - Institute of Cognitive Psychology; [2] University of Paris Sud - UFR STAPS

Abstract

Implicit mental rotation tasks involving body parts have been shown to reflect biomechanical properties of the corresponding motor performance. This finding not only buttresses the participation of motor processes in egocentric mental rotations as it raises additional issues. To what extent does motor imagery underlie mental rotation outside the realm of body segments? To what degree can it be instrumental in understanding strategies put forward by motor disabled persons? This work compared 24 normal and 24 palsy subjects on two mental rotation tasks concerning realistic hands photos (handedness judgments) and 3-D letters (normal/mirror judgments). Main outcomes go as follows: both groups exhibit a similar pattern of constraints on the hands task; longer response times are displayed by the palsy group on both tasks; overall, an effect of severe motility impairment on mental rotation of nonbody objects seems warranted, translating primarily in a lowered speed of letter rotation.

On the follow up of pioneer work by Shepard and Metzler (1971), employing pairs of drawings of misaligned 3-D objects, the paradigm of mental rotation was quickly extended, through some procedural variants, to different sorts of stimuli - letters, digits, random shapes, familiar objects. Exception made to peculiarities of the RT function (such as the upward curvilinear trend around 180° for alpha-numeric characters), the same major results were consistently obtained: a close to linear increase in RT as a function of angular distance, together with a clear choice for the shortest paths. A different extension requiring left-right judgments of body segments, which actually started soon after the early days (Cooper & Shepard, 1975), produced however distinct outcomes: just as mental rotation of shapes seemed to preserve analogue properties of the physical medium, mental rotation of body parts, e.g. hands or feet, seemed to inherit the mechanical constraints of real body movements, expressing as a preference for plausible biomechanical paths over shorter but implausible ones.

L. Parsons (1987, 1994) systematically pursued this line of inquiry in the framework of a general processing model of the mental simulation of one's action. Calling upon phenomenological (e.g., the reporting of kinaesthetic sensations while engaging in left-right judgments over body parts) behavioural (e.g., strong correlations between the time required to issue a hand laterality judgment, to actually move one's hand into the orientation of the stimulus, and to simply imagine the corresponding movement), and neuropsychological evidence (e.g., substantial overlap of activation areas underlying laterality judgments and actual movements), he made a compelling case for an intimate link between action and mental rotation through motor imagery. The ability of chronometric rotation paradigms to express dynamical constraints of motor behaviour was actually shown to go as far as to reflect transitional peripheral factors such as pain affecting a specific limb (Schwoebel at al., 2001), and has subsequently been exploited in studies addressing motor-impaired populations - e.g., Parkinsonic patients (Dominey, 1995) amputees and people congenitally deprived of one hand (Funk & Brugger, 2002) and hemiplegic stroke patients (Johnson-Frey, 2004).

Currently under debate is the functional relation between motor and visual imagery. A modularity thesis with roots that go back to A. Binet pretends that they are dissociated, drawing on different resources (Sirigu & Duhamel, 2001). A different stance is advocated, for instance, by Wexler et al. (1998), envisaging mental rotation as a covert simulation of *motor rotation* – the first natural reaction to be expected from a subject striving to align two Shepard-Metzler objects. Neurological

evidence regarding this contention is mixed, with studies supporting a general involvement of motor areas of the brain (e.g., Cohen et al., 1996), others restricting it to specific rotation tasks, such as those involving hands or tools (e.g., Kosslyn et al., 1998). The picture is further complicated by factors such as acknowledged individual differences in rotation strategies (e.g., Seurinck et al., 2004), the strategic role played by instructions (Sirigu & Duhamel, 2001) or the concrete nature of the tasks (Kosslyn et al., 1998), which can all impact decisively on the way visual an motor imagery are intermixed.
In this study we investigated the way subjects suffering from cerebral palsy – a congenital chronic motor disorder that affects mainly locomotor, gesture and postural control (Zabalia, 2002) – perform mental rotation tasks concerning hands and letters. Directly addressed, by means of comparison with a sample of controls, are thus the conveyed constraints and strategies implemented by people who could never actually perform the implicitly requested hand movements. However, through comparison of cerebral palsy subjects performance on the “handedness of hands” and on the “normal-mirror” judgements of 3D letters, the relationships between motor and visual imagery are also indirectly at stake (i.e., how does severe motility constraints impacts on a chiefly visuospatial imagery task?)

Method

Subjects: 24 cerebral palsy subjects suffering from severe congenital handicap of upper-limbs (age range: 16-35 yrs) and 24 control subjects (18-36 yrs) took part in the experiment. All palsy subjects were free from cognitive deficits, had no neurological record of parietal injury and were shown to perform at a normal level on an adapted handedness test (exclusion criteria). It was ensured that all palsy subjects could make effective use of mouse buttons as response keys.

Design and procedure: All subjects performed two tasks, a “letter rotation” and a “hands rotation” task. Half the subjects started with the former, the other half with the later. The “hands rotation” task was actually composed of 3 separate trial blocks corresponding to three different viewpoints taken over the hand. Subjects were run through six different orderings of the blocks, according to a double latin square balanced for carry-over effects (4 subjects per sequence). Procedure was the same on both tasks. Subjects sat at approximately 60 cm of a computer display, hands resting palm down near the mouse. Stimuli were randomly presented, with no time limit. The “letters task” required subjects to issue a “normal-mirror” judgment as quickly as possible by pressing the corresponding mouse buttons (left or right). The handedness task demanded a laterality (left-right) judgment over presented hands, again expressed by quickly pressing a mouse button (“left”- left, “right”- right).

Stimuli: *Letters task*. 3-D textured letters (F and R), 50% normal and 50% mirror-reversed, created with 3DStudioMax and subsequently rotated at successive 30° steps, both in the rightward and leftward direction (thus, between 0° and 180°).
Hands task. Stimuli were realistic photos of human hands. Three viewpoints were selected from among the ones used in Parsons, 1994 (up-right, palm down, fingers-front); the reference photos (illustrated below) were then rotated at steps of 30º between 0º and 180º, either away from the body medial-saggittal plan (lateral trajectory), either towards it (medial trajectory).

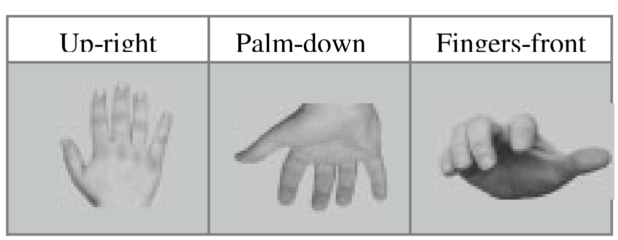


Figure 1. Viewpoints 1, 2, 3 - Reference photos (0º)

Results

I. Letters.

Figure 2 contrasts graphically groups' performances regarding latencies (A) and error rates (B). Both dependent variables present a "rotation effect" as a function of angular distance. While no substantive differences are apparent between groups on the error data, quite distinct "intercepts" and "linear slopes" (mean palsy: 17.26 *vs.* mean control: 6.58) are displayed by each group regarding RT (only RT from correct responses was considered for data analysis; on ground that palsy subjects surpassed the accuracy of controls when mirrored characters were separately considered, both kinds of letters were allowed to contribute to RT analysis).

A repeated measures ANOVA performed on the RT data revealed a highly significant effect for "Rotation"(p = .000), obeying furthermore a significant linear trend assessed by polynomial contrasts (replicated in separate ANOVAs for each group: palsy: p = .011; control: p = .012), as well as a highly significant effect for "Group" (p = .003). Non-parametric comparisons ($U_{M\text{-}W}$) between the distributions of "slopes" and intercepts in each group (limited to subjects providing a significant "linear fit" for RT data) were shown significant for "slope" (p = .001) and marginally significant for "ordinate" (p = .057).

Palsy subjects thus exhibited accuracy levels comparable to controls. They nevertheless required more time to perform the task. Part of this time increase, reflected on the heightened ordinate, can be accounted by a "response factor" favouring controls. More interestingly, however, another part concerns the differences in "slope", which indexes the speed of mental rotation. The major substantive difference between groups is thus a lowered rate of mental rotation in the palsy group.

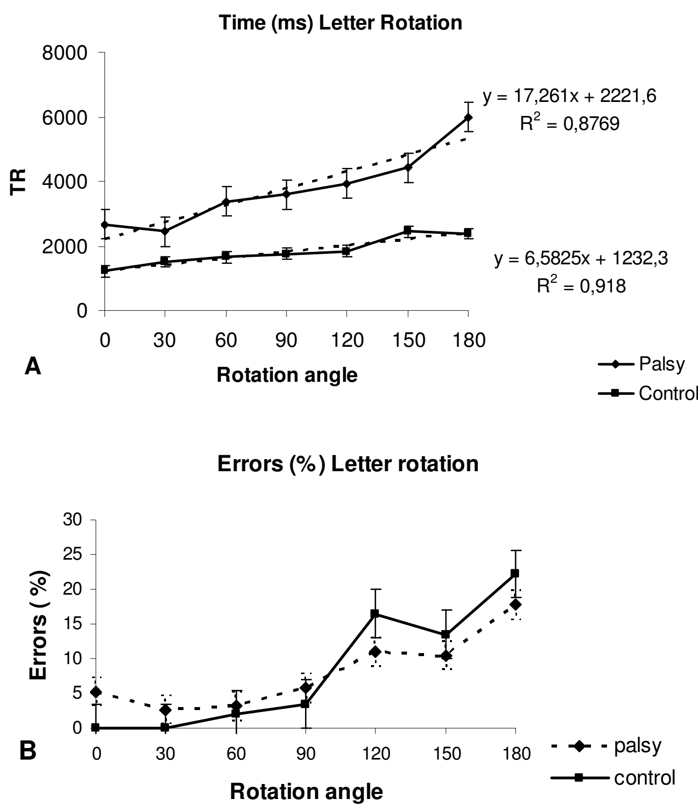


Figure 2. Letter rotation – Palsy and Control groups: (2**A**) RT (ms) ; (2**B**) % error

II. Hands

Figure 3 depicts the differences in RT (correct answers only) between *Lateral* (away from the body's medial-saggittal plan) and *Medial* (towards the plan) trajectories as a function of rotation angle, for each of the 3 viewpoints under consideration. Given that lateral trajectories are generally associated with increasing biomechanical constraints and more awkward hand postures, the net RT difference between Lateral and Medial paths of rotation (hereafter called DIFF) was taken as an index of the effects of dynamic constraints.

The first major feature to be noticed is the striking resemblance of the patterns obtained. Modulation of DIFF by "Viewpoint" (VP) follows the same rules in palsy and control groups: relative exemption from motor constraints on VP 1 ("DIFFs" floating around the zero reference line); positive DIFFs peaking around 90º on VP 2; reversed pattern of "DIFFs" (increased constraints associated with the medial trajectory) peaking at 120º on VP 3. As could be expected from this, medial and lateral RT patterns for each VP (not presented) were also structurally similar in both groups, the same happening on the whole with error (%) patterns (not presented).

Despite the similarities, important quantitative differences are apparent, with higher DIFFs in the palsy group. Statistical comparisons between groups on each VP (carried out through repeated measures ANOVA with "Group" as a between-subjects factor and "rotation" as a within-factor) were shown significant for VPs 2 and 3 (respectively: $p = .04$, $p = .044$). None of the Viewpoints showed a significant "Rotation*Group" interaction, which is consistent with the observed pattern similarity.

Likewise, RT differences between groups (favouring controls) on Lateral (L) and Medial (M) trajectories in each VP were shown significant (exception for Lateral path on VP 3). This finding is not entailed by the previous ones concerning DIFFs. While the former signal an increased effect of biomechanical constraints on palsy subject's implicit rotation, differences in L and M trajectories reflect general enlarged latencies of response.

Finally, not only the M and L latency patterns (not presented) accord well between groups as they are in overall agreement with the ones reported by Parsons (1994), as well as with subjective ratings of difficulty (scale format: 1-10) of actually produced hand postures independently obtained from a sample of eight normal subjects (not presented).

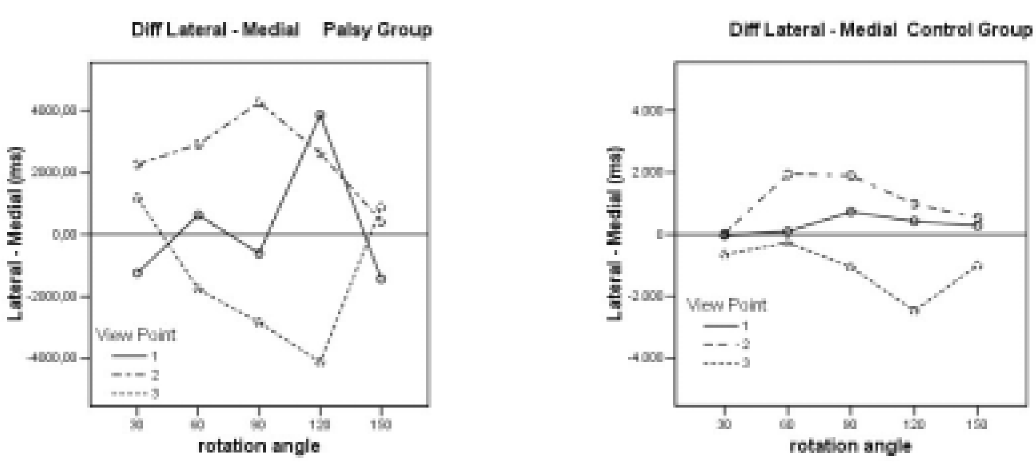


Figure 3. Biomechanical constraints (DIFF: RT lateral – RT medial) for Viewpoints 1,2 and 3

The pattern of correlations between "linear slope" in the letters task and the time required to issue the "left-right" judgments over hands provided an additional source of differences between groups (again, only subjects giving a significant linear fit contributed to the analysis). Whereas both groups presented significant correlations at several rotation angles in Viewpoint 1, only the palsy group went on displaying strong correlations at specific hand postures across Viewpoints 2 and 3. As expected from the independence of DIFFs regarding overall RTs, no significant correlations were observed between "linear slopes" and DIFFs in any group or viewpoint.

Discussion

As a general outcome, based on the closely resembling DIFF patterns and their further convergence with subjective ratings of difficulty, the same "implicit phenomenology" of motor constraints seems to be at work in both palsy and control subjects. Given that no actual motor counterpart of equivalent hands rotation is available to the palsy group, this raises the issue of a possible role of"vicariant motility". Mirror neurons, enabling a direct matching of observed and performed action (Buccino et al., 2004), could offer a plausible underlying candidate mechanism.

The clear expression of biomechanical constraints, indexed by DIFFs, in the palsy group, stands convincingly for the use of a "motor imagery" strategy by these subjects when performing the "hands task". However, the fact that significant correlations with "slope values" (obtained in the "letters task") are found in the palsy group all across viewpoints suggests that a mix of visual and motor strategies may have been put to use by them - contrasting with a more exclusive motor strategy of the controls on Viewpoints 2 and 3. The correlation patterns for Viewpoint 1 helps strengthening this interpretation: being the less dependent on wrist rotation and the less obviously 3-D, thence more prone to a visuospatial rotation as a nonbody object, it has shown widespread significant correlations with "linear slope" in either group. Consistently with this, DIFF values evolved around "zero" on VP 1 in both groups.

The heavy motor drawbacks affecting palsy subjects do seem to extend its effects beyond the hands task to the chiefly visuospatial letters rotation task. This effect expresses primarily as a significantly lowered speed of mental rotation (regarding controls). While from the standpoint of errors this task has not proved more difficult to palsy subjects than to controls (contrary to the "hands task"), the rate at which the assumed mental rotation took place differs considerably between groups. Dissociation between processing of the difficulty information and the process of mental rotation itself has been referred in the literature concerning mirror-normal characters discrimination, in connection with ERP effects of mental rotation (Heil, 2002). Also, processing models of imagery do contemplate for a distinct process of mentally rotating patterns in images (Mast et al., 2003). It is thus worth-while to envisage a specific effect of severe motor impairment on the speed of mental rotation.

On the other hand, that motor strategies can be covertly transferred to imaginal transformations of objects other than body parts has already been demonstrated in controlled experimental settings (Wraga et al., 2003). This allows speculating that visual mental rotation may arise as an easy expedient way, eventually rendered autonomous, of performing what started as a motor rotation strategy. On the whole, our findings do suggest that motility can be of importance for mental rotation generally, and not just for egocentric rotations of body segments.

Moreover, the employment of motor imagery strategies by the palsy group, and the shared phenomenology of the space of biomechanical constraints with the control subjects, pleads for the potential usefulness of imagined motor practice as a means to improve actual motor performance in palsy subjects, namely in the framework of ergonomic adaptation to tools and functional prosthesis.

References

Buccino G, Binkofski F. & Riggio, L. (2004). The mirror neuron system and action recognition. *Brain and Language*, **89**, 370-376.

Cohen, M., Kosslyn, S., Breiter, H., DiGirolamo, G., Thomson, W., Anderson, S. Bookheimer, B. & Belliveau, J. (1996). Changes in Cortical Activity During mental Rotation: A mapping study using functional magnetic resonance imaging. *Brain*, **119**, 89 – 1000.

Cooper, L., & Shepard, R. (1975). Mental transformations in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance,* **104**(1), 48-56.

Dominey, P., Decety J., Broussolle, E., Chabot, G. & Jeannerod, M. (1995). Motor imagery of a lateralized sequential task is asymmetrically slowed in hemi-Parkinson's patients. *Neuropsychologia*, **33**, 727-741.

Funk, M. & Brugger, P. (2002). Visual recognition of hands by persons born with only one hand. *Cortex*, **38**, 860-863.

Heil, M. (2002). The functional significance of ERP effects during mental rotation. *Psychophysiology,* **39***,* 535-545.

Johnson-Frey, S. (2004). Stimulating through simulation? Motor Imagery and functional reorganization in hemiplegic stroke patients. *Brain and cognition,* **55**, 328 – 331.

Kosslyn, S. M., DiGirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: neural mechanisms revealed by positron emission tomography. *Psychophysiology*, **35**, 151-161.

Mast, F., Ganis, G., Christie, S. & Kosslyn, S. M. (2003). Four types of visual mental imagery processing in upright and tilted observers. *Cognitive Brain Research*, **17**, 238-247.

Parsons, L.M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, **19**, 178-241.

Parsons, L.M. (1994). Temporal and Kinematic properties of Motor Behavior Reflected in Mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance,* **20** (4) 709 – 703.

Schwoebel, J., Friedman R., Duda, N. & Coslett H. (2001). Pain and the body schema: Evidence for peripheral effects on mental representations of movement. *Brain*, **124**, 2098-2104.

Seurink, R. Vingerhoets, F. P. & Achten, E. (2004). Does egocentric mental rotation elicit sex differences? *Neuroimage*, **23**, 1440-1449.

Shepard, R., N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, **171**, 701-703

Sirigu, A. & Duhamel, J. (2001). Motor and visual imagery as two complementary but neurally dissociable mental processes. *Journal of Cognitive Neuroscience*, **13** (7), 910-919.

Wexler, M., Kosslyn, S.M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, **68**, 77-94.

Wraga, M., Thompson, W. L., Alpert, N. M. & Kosslyn, S. M. (2003). Implicit transfer of motor strategies in mental rotation. *Brain and Cognition*, **52**, 135-143.

Zabalia, M. (2002). Action and mental imagery in children. *Année Psychologique,* **102** (3), 409-422.

TEST OF LINEARITY OF THE RESPONSE FUNCTION FOR RATINGS OF PERCEIVED AREA

Sergio Cesare Masin
University of Padua, Italy
scm@unipd.it

Abstract

The linearity of the response function for ratings of perceived area was tested. The results show that this function is linear if the response function for magnitude estimates of perceived length is linear. A problem for future research is pinpointed.

The response function relates measures of sensory intensity to values of sensory intensity. There is evidence that this function is linear for ratings of sensory intensity (Anderson, 1981, 1982, 1996, pp. 94-96; Curtis & Fox, 1969). The present study explored whether the response function was linear for ratings of perceived area (hereafter called area) of surfaces presented frontally.

The area of a rectangle presented frontally is

$$\alpha_R = \omega\, \eta \tag{1}$$

with ω the perceived width and η the perceived height of the rectangle.

The psychophysical function relates measures of sensory intensity to measures of physical intensity. This function is linear when it is obtained from magnitude estimates of perceived length (hereafter called length) of lines presented frontally (Baird & Vernon, 1965; Bogartz, 1979; Ekman & Junge, 1961; Fagot, 1982; Hartley, 1977, 1981; Irvin & Verrillo, 1979; Kerst & Howard, 1983; Masin & Vidotto, 1983; Pitz, 1965; Reese, Reese, Volkmann, & Corbin, 1953; Schiffman, 1965; Stevens & Galanter, 1957; Stevens & Guirao, 1963; Svenson & Åkesson, 1966; Teghtsoonian, 1965; Teghtsoonian & Beckwith, 1976; Teghtsoonian & Teghtsoonian, 1971; Verrillo, 1981, 1982, 1983; Zwislocki & Goodman, 1980). Assuming that the response function for magnitude estimates of length is linear, the linearity of the psychophysical function for length implies that

$$\omega = k_0\, w + k_1 \tag{2}$$

and

$$\eta = k_0\, h + k_1 \tag{3}$$

with w the measure of the physical width and h the measure of the physical height of the rectangle and with k_0 and k_1 unknown parameters.

Equations 1–3 yield

$$\alpha_R = k_0^2\, w\, h + k_0\, k_1\, (w + h) + k_1^2\ . \tag{4}$$

Let us assume that the response function for the rating R_R of α_R is

$$R_R = c_0\, \alpha_R + c_1 \tag{5}$$

with c_0 and c_1 unknown parameters.

Equations 4 and 5 yield

$$R_R = c_0\, k_0^2\, w\, h + c_0\, k_0\, k_1\, (w + h) + c_0\, k_1^2 + c_1\, . \tag{6}$$

Equation 6 implies the prediction that R_R varies linearly with w when h is fixed. In the experiment reported below, subjects rated the area of 13 rectangles with h fixed at 21 cm and with w varying in steps of 1.5 cm from 3 to 21 cm. The linearity of the response function for ratings of area (Equation 5) was tested by testing whether R_R varied linearly with w.

Note that this test is based on the assumption that the response function for magnitude estimates of length is linear. If the prediction that R_R varies linearly with w is verified, one concludes that the response function for ratings of area is truly linear if the response function for magnitude estimates of length is truly linear.

In the experiment reported below, subjects were asked to rate the area of 13 disks of different area. Each physical area of disks was equal to the physical area of one of the rectangles used to test Equation 5. It may easily be shown that the area of these disks was

$$\alpha_D = k_0^2\, w\, h + 2\, k_0\, k_1 \sqrt{\pi\, w\, h} + \pi\, k_1^2\, . \tag{7}$$

Consequently the rating of α_D was

$$R_D = c_0\, k_0^2\, w\, h + 2\, c_0\, k_0\, k_1 \sqrt{\pi\, w\, h} + \pi\, c_0\, k_1^2 + c_1\, . \tag{8}$$

I have calculated that the root-mean-square deviation of mean ratings of rectangle area obtained by Anderson and Cuneo (1978) from corresponding R_Rs predicted by Equation 6 is minimized when $k_1 = 0$. For h fixed, Equation 8 shows that R_D varies linearly with w if $k_1 = 0$ and varies nonlinearly with w if $k_1 \neq 0$. The possibility that $k_1 = 0$ was tested by testing whether R_D varied linearly with w.

To appraise sensitivity of ratings to nonlinearity subjects were asked to rate the length of 13 horizontal lines of different length. Each physical length of lines equaled the physical diameter of one of the disks used to test whether $k_1 = 0$. Since

$$\alpha_D = \tfrac{1}{4}\, \pi\, \delta^2 \tag{9}$$

with δ the perceived diameter of the disk, it must be that

$$\alpha_D = \tfrac{1}{4}\, \pi\, (k_0^2\, d^2 + 2\, k_0\, k_1\, d + k_1^2)\, . \tag{10}$$

with d the measure of the physical diameter of the disk. Equation 10 shows that ratings of line length must vary nonlinearly with w. Sensitivity to nonlinearity was appraised by testing this implication.

Method

Subjects

Nineteen university students participated in the experiment as subjects.

Stimuli

Experimental stimuli were achromatic rectangles, disks, or horizontal lines each with luminance of 5 cd/m^2 located in the middle of a 83 × 60 cm 25 cd/m^2 achromatic rectangular background presented frontally in the middle of the screen of a horizontal NRC PlasmaSync 50MP2 plasma monitor controlled by a Power Maintosh 7200/90 computer. Viewing distance was 270 cm. The experimental room was illuminated only by the monitor screen.

There were thirteen rectangles all with height 21 cm and with width varying in steps of 1.5 cm from 3 to 21 cm. For each rectangle there was one disk with physical area equal to that of the rectangle. For each disk there was one horizontal 1 pixel wide line with physical length equal to the physical diameter of the disk. Stimuli were shown twice randomly. To compensate for orientation each rectangle was shown once horizontally and once vertically.

Two 5 cd/m^2 achromatic standard stimuli were presented in the middle of the screen before each experimental stimulus. For rectangles or disks the standard stimuli were two squares with horizontally aligned centers, one with side length of 4 cm and one with side length of 30 cm. For horizontal lines the standard stimuli were two collinear 1 pixel wide horizontal lines, one with length of 4 cm and one with length of 30 cm. The width of the gap between the standard stimuli was 16 cm. The standard stimuli appeared for 1 sec, randomly in one of the two possible relative positions. The time between the offset of the standard stimuli and the onset of the corresponding experimental stimulus was of 1 sec. The experimental stimulus disappeared when the experimenter typed the response of the subject. Standard stimuli appeared 1 sec after this response was typed.

Procedure

The following instructions were displayed on the monitor screen and were read and commented when necessary by the experimenter: “In this experiment, you will be shown squares, rectangles, disks, and horizontal lines, one at a time. You are asked to rate how much the areas of the squares, of the rectangles, and of the disks are large and how much the lines are long. Ratings are to be expressed using the integer numbers from 10 to 100. The following are the two standard stimuli presented before each square, each rectangle, and each disk (the standard stimuli made of squares were presented once, with relative position selected randomly). The area of the smallest square is equal to 10 and the area of largest square is equal to 100. The following are the two standard stimuli presented before each line (the standard stimuli made of lines were presented once, with relative position selected randomly). The length of the shortest line is equal to 10 and the length of longest line is equal to 100. Each number assigned to the squares, rectangles, or disks must be in proportion to their area—the larger the area the larger the number—considering that the area of the smaller standard is 10 and that the area of the larger standard is 100. Each number assigned to the lines must be in proportion to their length—the longer the line the larger the number—considering that the length of the shorter standard is 10 and that the length of the longer standard is 100.” A large response range and two standard stimuli, one much smaller than the smallest and one much larger than the largest experimental stimulus, were used to minimize biases (Foley, Cross, Foley, & Fox, 1983; Marks 1968; Parducci 1982; Parducci & Wedell 1986). Integers for ratings were restricted in the range 10–100 to avoid the bias due to the preference of individuals for digits (Baird & Noma, 1978, p. 109).

Main results

In Figure 1, the left and central diagrams show, respectively, the mean ratings of rectangle area and of disk area as a function of rectangle width, with width defined as above. In the right diagram, the larger dots show mean ratings of length of horizontal lines as a function of rectangle width, while the smaller dots show these mean ratings as a function of disk diameter. For each stimulus, the individual score for each subject was the mean of the two ratings the subject assigned to the stimulus.

The results for rectangles agree with previous findings (Anderson & Weiss, 1971). A least-squares straight line fits mean ratings of area of rectangles as a function of rectangle width. The linear trend was significant [$F(1,18) = 359$, $p < 0.0005$] and the quadratic trend was not significant [$F(1,18) = 1.97$]. These results confirm Equation 5.

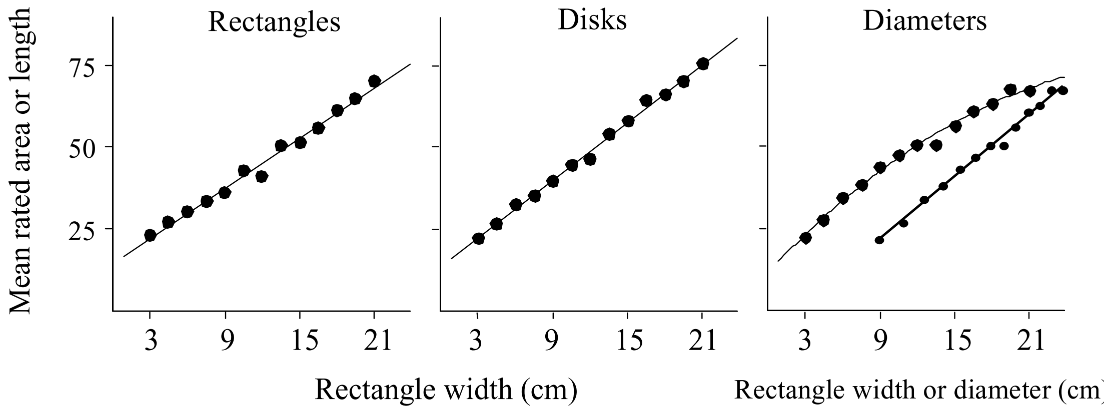


Figure 1. Mean rated area of rectangles and of disks as a function of rectangle width and mean rated length of lines as a function of rectangle width (larger dots) or of disk diameter (smaller dots).

A least-squares straight line fits mean ratings of area of disks as a function of rectangle width. The linear trend was significant [$F(1,18) = 544$, $p < 0.0005$] and the quadratic trend was not significant [$F(1,18) = 0.002$]. These results confirm that $k_1 = 0$.

In the right diagram a least-squares straight line fits mean ratings of horizontal line length as a function of disk diameter (smaller dots). This straight line shows that the psychophysical function had an exponent of 1. Stevens and Galanter (1957) found that ratings produced a psychophysical function for length with exponent 0.69 (Ward, 1974) contributing influentially to the negative view that ratings were biased. However, Stevens and Galanter (1957) used ratings without following the methodological precautions that are known today to be necessary to minimize context effects.

A least-squares parabolic arc fits the mean ratings of horizontal line length as a function of rectangle width (larger dots). The linear and quadratic trends were significant [Fs(1,18) = 302 and 19.6, respectively, $p < 0.0005$]. These results show that ratings were sensitive to nonlinearity and thus confirm Equation 5. When squared individual scores rather than individual scores were used for the statistical analyses, the linear trend was significant [$F(1,18) = 121$, $p < 0.0005$] but the quadratic trend was no longer significant [$F(1,18) = 0.41$] in conformity with the fact that disk area varied linearly with the square of the diameter (Equation 9). These results confirm that $k_1 = 0$.

Serendipetous results

A 2 (rectangle vs. disk) × 13 (rectangle width) analysis of variance showed that mean ratings of disk area were significantly higher than mean ratings of rectangle area [$F(1,18) = 11.9$, $p < 0.005$]. The mean ratings of disk area and rectangle area progressively diverged as physical area increased. The interaction was marginally significant [$F(12,216) = 1.73$, $p = 0.06$].

This finding that the shape of stimuli had an effect on rated area has no relevant implication for the line of reasoning of the present study.

It is undetermined whether the effect of shape was perceptual, mnemonic, or both. On one hand it could be that an illusory change in ω or η increasing with area caused this effect. On the other hand the following results indicate that the effect of shape could have been a memory rather then a perceptual effect. In two carefully executed experiments, Bolton (1897; see also Anastasi, 1936) had 25 subjects either select a square so that its area matched the area of a standard disk, or select a disk so that its area matched the area of a standard square. Essentially the results showed that the areas of the surfaces matched when the corresponding physical areas matched. These results indicate that the overestimation of the area of disks found in the present study could have been due to the successive comparison of experimental stimuli with remembered standard stimuli.

Conclusion

The assumption that the response function for magnitude estimates of length is linear implies Equations 2 and 3. Functional measurement and the bisection method confirm this assumption since they confirm these equations (Anderson, 1974, 1977). The present results show that the response function for ratings of area is linear if the response function for magnitude estimates of length is linear.

This conclusion leads to a problem that deserves investigation.

Since the psychophysical function for length obtained by magnitude estimation is linear and the results in Figure 1 show that the psychophysical function for length obtained by ratings is linear, magnitude estimation and ratings should both involve a linear response function for length.

The psychophysical function for area obtained by magnitude estimation is nonlinear with exponent of about 0.75 (Baird, 1970; Da Silva, Marquez, & Ruiz, 1987; Rule & Markley, 1971; Teghtsoonian, 1965; Teghtsoonian & Teghtsoonian, 1983) and the results in Figure 1 show that the psychophysical function for area obtained by ratings is linear. Thus, since the present results show that the response function for ratings of area is linear (if the response function for magnitude estimates of length is linear) it should be that the response function for magnitude estimates of area is nonlinear.

The problem is, why should the response function for magnitude estimates of area be nonlinear?

References

Anastasi, A. (1936). The estimation of area. *Journal of General Psychology*, **14**, 201-225.

Anderson, N. H. (1974). Cross-task validation of functional measurement using judgments of total magnitude. *Journal of Experimental Psychology*, **102**, 226-233.

Anderson, N. H. (1977). Failure of additivity in bisection of length. *Perception & Psychophysics*, **22**, 213-222.

Anderson, N. H. (1981). *Foundations of Information Integration Theory*. New York: Academic Press.

Anderson, N. H. (1982). *Methods of Information Integration Theory*. New York: Academic Press.

Anderson, N. H. (1996). *A functional theory of cognition*. Mahwah, NJ: Erlbaum.

Anderson, N. H., & Cuneo, D. O. (1978). The height + width rule in children's judgments of quantity. *Journal of Experimental Psychology: General*, **107**, 335-378.

Anderson, N. H., & Weiss, D. J. (1971). Test of a multiplying model for estimated area of rectangles. *American Journal of Psychology*, **84**, 543-548.

Baird, J. C. (1970). *Psychophysical analysis of visual space*. New York: Pergamon Press.

Baird, J. C., & Noma, E. (1978). *Fundamentals of scaling and psychophysics*. New York: Wiley.

Baird, J. C., & Vernon, C. W. (1965). Binary estimation. *Psychonomic Science*, **3**, 469-470.

Bogartz, R. S. (1979). Line ratio judgments yield subjective lengths proportional to physical lengths: reanalysis of Engen's data. *Perception & Psychophysics*, **26**, 247-249.

Bolton, F. A. (1897). A contribution to the study of illusions. *American Journal of Psychology*, **9**, 167-182.

Curtis, D. W., & Fox, B. E. (1969). Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics*, **5**, 89-93.

Da Silva, J. A., Marquez, S. L., & Ruiz, E. M. (1987). Subjective differences in exponents of psychophysical power functions for inferred, remembered, and perceived area. *Bulletin of the Psychonomic Society*, **25**, 191-194.

Ekman, G., & Junge, K. (1961). Psychophysical relations in visual perception of length, area, and volume. *Scandinavian Journal of Psychology*, **2**, 1-10.

Fagot, R. F. (1982). Constructability and the power function for line judgments. *Perception & Psychophysics*, **31**, 392-394.

Foley, H. J., Cross, D. V., Foley, A., & Fox, R. (1983). Stimulus range, number of categories, and the “virtual exponent”. *Perception & Psychophysics*, **34**, 505-512.

Hartley, A. A. (1977). Mental measurement in the magnitude estimation of length. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 622-628.

Hartley, A. A. (1981). Mental measurement of length: the role of the standard. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 309-317.

Irvin, G., & Verrillo, R. T. (1979). Absolute estimation of line length and afterimage as a function of viewing distance. *Sensory Processes*, **3**, 275-285.

Kerst, S. M., & Howard, J. H. (1983). Mental processes in magnitude estimation of length and loudness. *Bulletin of the Psychonomic Society*, **21**, 141-144.

Marks, L. E. (1968). Stimulus-range, number of categories, and form of the category-scale. *American Journal of Psychology*, **81**, 467-479.

Masin, S. C., & Vidotto, G. (1983). A magnitude estimation study of the inverted-T illusion. *Perception & Psychophysics*, **33**, 582-584.

Parducci, A. (1982). Category ratings: still more contextual effects! In B. Wegener (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, NJ: Erlbaum. (pp. 89-105).

Parducci, A., & Wedell, D. H. (1986). The category effect with rating scales: number of categories, number of stimuli, and method of presentation. *Journal of Experimental Psychology: Human Perception and Performance*, **12**, 496-516.

Pitz, G. F. (1965). Magnitude scales of line length. *Psychonomic Science*, **2**, 213-214.

Reese, E. P., Reese, T. W., Volkmann, J., & Corbin, H. H. (Eds.) (1953). *Psychophysical research, summary report 1946-1952* (Technical report No. SDC 131-1-15). South Handley, MA: U.S.N. Special Devices Center.

Rule, S. J., & Markley, R. P. (1971). Subjective differences in cross-modality matching. *Perception & Psychophysics*, **9**, 115-117.

Schiffman, H. R. (1965). Size estimation and the size of the measuring unit. *Psychonomic Science*, **3**, 479-480.

Stevens, S. S., & Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, **54**, 377-411.

Stevens, S. S., & Guirao, M. (1963). Subjective scaling of length and area and the matching of length to loudness and brightness. *Journal of Experimental Psychology*, **66**, 177-186.

Svenson, O., & Åkesson, C. A. (1966). *Fractional and multiple estimates in ratio scaling* (Report No. 202). Stockholm: Psychological Laboratory, University of Stockholm.

Teghtsoonian, M. (1965). The judgment of size. *American Journal of Psychology*, **78**, 393-402.

Teghtsoonian, M., & Beckwith, J. B. (1976). Children’s size judgments when size and distance vary: is there a developmental trend to overconstancy? *Journal of Experimental Child Psychology*, **22**, 23-39.

Teghtsoonian, M., & Teghtsoonian, R. (1971). How repeatable are Stevens’s power law exponents for individual subjects? *Perception & Psychophysics*, **10**, 147-149.

Teghtsoonian, M., & Teghtsoonian, R. (1983). Consistency of individual exponents in cross-modal matching. *Perception & Psychophysics*, **33**, 203-214.

Verrillo, R. T. (1981). Absolute estimation of line length in three age groups. *Journal of Gerontology*, **36**, 625-627.

Verrillo, R. T. (1982). Absolute estimation of line length as a function of sex. *Bulletin of the Psychonomic Society*, **19**, 334-335.

Verrillo, R. T. (1983). Stability of line-length estimates using the method of absolute magnitude estimation. *Perception & Psychophysics*, **33**, 261-265.

Ward, L. M. (1974). Power functions for category judgments of duration and line length. *Perceptual and Motor Skills*, **38**, 1182.

Zwislocki, J. J., & Goodman, D. A. (1980). Absolute scaling of sensory magnitudes: a validation. *Perception & Psychophysics*, **28**, 28-38.

LEVEL AND INTER-STIMULUS INTERVAL EFFECTS: TIME ORDER ERRORS IN VISUAL LENGTH DISCRIMINATION

David McGill and William M. Petrusic
Department of Psychology, Carleton University, Ottawa, Ontario, Canada, K1S 5B6
dgmcgill@hotmail.com, bpetrusi@ccs.carleton.ca

Abstract

Time order error (TOE) effects were found in two paired comparison experiments, which used line length stimuli. Participants [RM1]tended to exhibit greatest discriminative accuracy in the short-long order of stimulus presentation. Inter-stimulus interval (ISI) and magnitude of the comparison pair modulated this TOE effect. Singular attempts to explain the TOE in terms of sensory fatigue, peripheral or decisional biases, and assimilation-contrast were unsuccessful. Two broad theories of comparative judgment were evaluated. Hellström's (1985) differential sensation weighting theory was tested with a linear probability model. Link's (1992) wave theory was tested with a logistic regression model. The analyses suggested that TOE is not a product of bias, and showed that the differential sensation weighting model provides a reasonable account of the comparison process.

Psychologists have studied the TOE phenomenon since Gustav Theodor Fechner established psychophysics in the mid-nineteenth century. Fechner (1966) served as his own participant in a series of experiments involving the comparison of lifted weights. Using the method of constant stimuli, he observed that the second-lifted comparison weights in his experiments tended to be [RM2]overestimated relative to the first-lifted standard weights. Negative TOE occurs when second-presented stimuli are over-estimated relative to first-presented stimuli. Positive TOE occurs in the converse scenario. In either case, binary judgment is affected by the order in which two stimuli are compared.

Since Fechner's day, TOE has been further documented in comparative judgment experiments involving weights (Hellström, 1985), sound intensities (Ellis, 1973), temporal durations (Jamieson & Petrusic 1975), line lengths (Petrusic, Harrison & Baranski, 2004), area for geometric shapes (Inomata, 1959), electric shock intensity (Geertsma, 1958), and dot numerosity (Rambo & Johnson, 1964). An adequate theory of TOE must dauntingly cut across sensory domains to outline the fundamental processes underlying comparative judgment. Consequently, TOE is not an insular, unsolved mystery of psychophysics. Waylaid theories of presentation order effects have inexorably contributed to the broader topic of comparative judgment.

Fading trace theory

Köhler (as cited in Needham, 1934) theorized that each physical stimulus is internally represented by way of a memory trace. He further supposed that this trace increases immediately after stimulus presentation, quickly reaches asymptote, and then begins to fade. Köhler's trace strength is synonymous with subjective magnitude. A first-presented stimulus, being held in memory, should be overestimated if judgment is made while its trace is near the initial asymptote (producing positive TOE). Meanwhile, a first-presented stimulus should be underestimated if judgment is made when the trace is in a state of decline (producing negative TOE). Fading trace theory readily explains findings of positive TOE when an ISI is brief. It also explains findings of increasingly negative TOE as ISI lengthens. Nonetheless, fading trace theory is contradicted by findings that positive TOE increases with ISI. The theory is also insufficient because it fails to explain how TOE is affected by magnitude of a stimulus pair.

Simple response bias

A simple response bias exists when a participant prefers certain modes of response. In the comparative judgment paradigm, there are typically two modes of response (e.g., respond "less" or "greater"). If TOE were the product of a participant's preference for one of two possible responses, then changing instructions from "determine if the second stimulus is greater" to "determine if the second stimulus is less" would reverse the direction of the error. Changing instructions in this manner has not been found to reverse TOE (cf., Hellström, 1977; Jamieson & Petrusic, 1975). On this basis, simple bias can be dismissed as a source of TOE.

Differential sensation weighting

In the Thurstonian tradition, Hellström's differential sensation weighting model (1985, 2000) holds that response proportions are a linear function of the subjective scale values for compared items. The model's testable postulate is that stimuli regress towards or away from a relational level (RL), depending upon how the stimulus-RL relationship is weighted. As such, differential sensation weighting does not involve the direct modification of subjective stimulus values. Rather, it is the difference between a stimulus and its RL that is weighted. Judgment is primarily influenced by two factors: i) RL values, and ii) the extent to which a stimulus[RM3] scale value assimilates towards[RM4] or contrasts with a RL. Hellström's RL is not constrained to be the geometric mean of a stimulus series; it can take on any value. The differential sensation weighting model is expressed as

$$D = k\ [s_1\psi_1 + (1 - s_1)\psi_{RL1}] - [s_2\ \psi_2 + (1 - s_2)\ \psi_{RL2}], \tag{1}$$

where D is subjective difference, s_1 and s_2 represent weighting coefficients, and ψ_1 and ψ_2 denote subjective scale values for the first and second stimuli. ψ_{RL1} and ψ_{RL2} reflect RLs for the first and second stimuli. The scale constant associated with the relevant psychometric function is k. When predicting response proportions for a binary comparative judgment task, the model becomes

$$z_{LG} = (ks_1/\sigma_c)\ \psi_1 - (k\ s_2/\sigma_c)\ \psi_2 + [- k\ s_2\ \psi_{RL2}\ - ks_1\ \psi\ _{RL1} + k(\psi_{RL1} - \psi\ _{RL2})/\sigma_c)], \tag{2}$$

where σ_c is the comparatal standard deviation for the distribution of subjective differences. Hellström has often found that $RL_1 = RL_2$, and this is assumed to be the case in subsequent discussions of the model.

Wave theory

Link's (1992) wave theory of comparison attributes TOE effects to decisional bias. Link proposed that each stimulus is internally represented by a continuous, electromagnetic wave. When two stimuli are compared, an accumulator continually sums the differences between wave amplitudes. The contents of this accumulator are compared to upper (greater) and lower (lesser) response thresholds. A response is made when the summed difference between wave amplitudes exceeds either response threshold. A decisional bias exists when one response is favoured at the expense of the alternative.

Link models the proportion of "second greater" responses in a binary discriminative judgment task as follows:

$$P_{LG} = \frac{e^{(\Delta S/S)T} - e^{(\Delta S/S)B}}{e^{(\Delta S/S)T} - e^{-(\Delta S/S)T}}. \tag{3}$$

P_{LG} represents the proportion of "second greater" judgments associated with a comparison pair. The base of the natural logarithm is represented by e (≈ 2.718). ΔS is the change needed to produce a just noticeable difference and S is the standard stimulus. T represents the response threshold, and B is a bias parameter. The response threshold, T, is not expected to change over ISI (Link, 1992, p. 198). However, ΔS, and bias may change. In Equation 3, S serves as the only variable that can be assessed prior to data analysis.

One may substitute for (ΔS/S)T if the method of paired comparisons is employed (Link, 1992, p. 246). In this case, the model becomes

$$P_{LG} = \frac{e^{[\ln(Sj) - \ln(ÓSi/N)]\,T} - e^{[\ln(Sj) - \ln(ÓSi/N)]\,B}}{e^{[\ln(Sj) - \ln(ÓSi/N)]\,T} - e^{-[\ln(Sj) - \ln(ÓSi/N)]\,T}}, \tag{4}$$

where S_j are the standard stimuli, and S_i are the N comparison stimuli for each standard.

Experiment 1

Method

Participants. Twenty undergraduate students participated in two 1-hour sessions to fulfill a course requirement.

Apparatus. SuperLab Pro V. 2.0, a software program, regulated instruction presentation, stimulus presentation, timing, event sequencing, and the recording of responses. This program was run on a Pentium II Intel NMX computer with 511 MB of RAM. Visual display was provided via a 19 inch ViewSonic PF790 monitor. Participant responses were issued via a Microsoft serial mouse.

Stimuli. The stimuli were black, horizontal, visual extents measuring 2 pixels in thickness. Three pairs of extents were used. These pairs were 50-51, 200-204, and 400-408 pixels in length. Visual angles for the shortest members of each pair, at a viewing distance of 40 cm, were 1.2°, 4.7°, and 9.3°.

Design. A 3 X 3 X 2 X 2 within-subjects design was used. There were three levels of stimulus pair (above described), three levels of ISI (500 ms, 3000 ms, and 6000 ms), two levels for presentation order (short-long and long-short), and two levels for instruction (shorter or longer [RM5]). Participants performed 12 repetitions of each condition.

Procedure. Each trial began with the appearance of a "SHORTER" or "LONGER" instruction (i.e. select the shorter, or longer line). These words appeared above the line presentation location, and preceded presentation by 1000 ms. The first comparison line remained on the screen for 2000 ms before disappearing. After the ISI, the second comparison line appeared, and remained on the screen until participants made a response by right- or left-clicking the mouse. This second line was presented in a location that was horizontally offset from the first line. The inter-trial interval was fixed at 2000 ms.

Results

Analysis of variance. A four-way within-subjects [RM6]analysis of variance was performed to evaluate the effects of pair length, presentation order, instruction type, and ISI using the 2 X arcsine transformed proportion of correct responses as the dependent measure. The pair X order X ISI interaction was significant, $F(4, 76) = 5.037$, $p < .01$, partial $\eta^2 = .21$.

Figure 1 shows that TOE became increasingly positive with ISI for the shortest line pair. It remained uniformly negative across ISI for the mid-range line pair. Finally, TOE became increasingly negative with ISI for the longest line pair.

Differential sensation weighting analyses. A multivariate regression was performed at each ISI for data that were collapsed across participants. In this regression, the standard normal deviate for the proportion of "longer" responses (z_{LG}) was a function of first [RM7]and second line lengths.

The differential sensation weighting model was evaluated in the context of two different psychophysical functions: Steven's power law and Fechner's law. If one presumes Steven's power law to be the psychophysical function, then Equation 2 is tested with the formula,

$$z_{LG} = a_1\, x_1^{\,b} + a_2\, x_2^{\,b} + a_0, \tag{5}$$

where x_1 and x_2 represent first and second stimuli. Meanwhile, b, a_1, a_2, and a_0 are fitted coefficients. The following formula is used to test the model that incorporates Fechner's law:

$$z_{LG} = a_1 \log[x_1] + a_2 \log[x_2] + a_0. \tag{6}$$

Equations 5 and 6 were both fit to the data. The increment in explained variance associated with using the power law instead of the logarithmic law did not approach significance, $F(1,15)$ =

0.00, p = 1.00. Thus, there was no benefit associated with adding an extra coefficient to the regression equation. Subsequent results pertain to Equation 6.

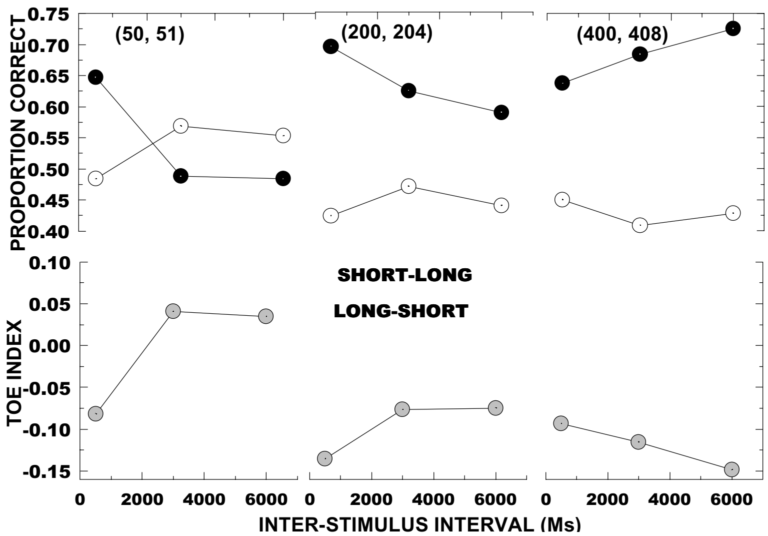


Figure 1. A three-way interaction between ISI, presentation order, and stimulus pair in terms of proportion correct. A two-way interaction between ISI and stimulus pair in terms of Jamieson and Petrusic's (1975) T_e index of TOE.

Since the data sets were particularly small, R_{press}^2 was chosen as a measure of association. This measure of association is always equal to, or lower than R^2. The difference between the two statistics increases as one progressively overfits the data.

The differential sensation weighting model significantly fit data at the 500 ms level, $F(2,3)$ = 10.43, $p < .05$, R = .94, $R_{press}^2 = .55$. It also fit significantly at the 3000 ms level [$F(2,3)$ = 152.85, $p <$ .01, R = .995, R_{press}^2 = .94], and at the 6000 ms level, $F(2,3)$ = 152.85, $p < .01$, R = .94, R_{press}^2 = .53. All coefficients were significant to the p =.01 level, except for a_0 at the 500 ms ISI, which did not approach significance. The calculated RL was 293 pixels and greatly exceeded the geometric mean of the stimulus series, which was 160 pixels. Weight ratios (s_1/s_2) for the 500 ms, 3000 ms, and 6000 ms ISIs were 0.996, 0.998, and 0.953, respectively. This latter result provides ambiguous support for the notion that RLs increasingly influence subjective scale values over time.

Wave theory analyses. Equation 4 was empirically fit to the data according to least squares criteria. The two parameters, T and B, were allowed to vary over ISI. Since there was only one comparison stimulus for each standard, ln ($ÓS_i$/N) simply became ln (S_i). The regression performed at the 500 ms level of ISI was significant, $F(1,4)$ = 21.97, $p < .01$, R = .92, R_{press}^2 = .65. Neither of the coefficients in this regression were statistically significant. The model poorly fit the data associated with the 3000 ms ISI [$F(1,4)$ = 1.11, p = .35, R = .47, $R_{press}^2 < 0$], and the 6000 ms ISI, $F(1,4)$ = 0.80, p = .42, R = .41, $R_{press}^2 < 0$.

Experiment 2

Method

Participants. Eighteen undergraduate students participated in two 1-hour sessions to fulfill a course requirement.

Apparatus. The equipment were identical to those used in the first experiment.

Stimuli. The stimuli were black, horizontal, visual extents measuring 2 pixels in thickness. Three pairs of extents were used. These pairs were 33-34, 200-206, and 400-412 pixels in length.

Design and procedure. The second experiment was identical to the first in terms of design and procedure. It was intended to provide systematic replication.

Results

Analysis of variance. The sphericity assumption had to be rejected for a pair X order X ISI interaction. Thus, the Huynh-Feldt correction to degrees of freedom was used when determining the three-way interaction to be significant, $F(3.34, 56.70) = 5.27$, $p < .01$, partial $\eta^2 = .24$. This interaction was remarkably similar to that which was obtained in the first experiment. Negative TOEs became increasingly positive with ISI for the shortest line pair. Positive TOEs became increasingly negative with ISI for the longest line pair. This interaction is plotted in Figure 2.

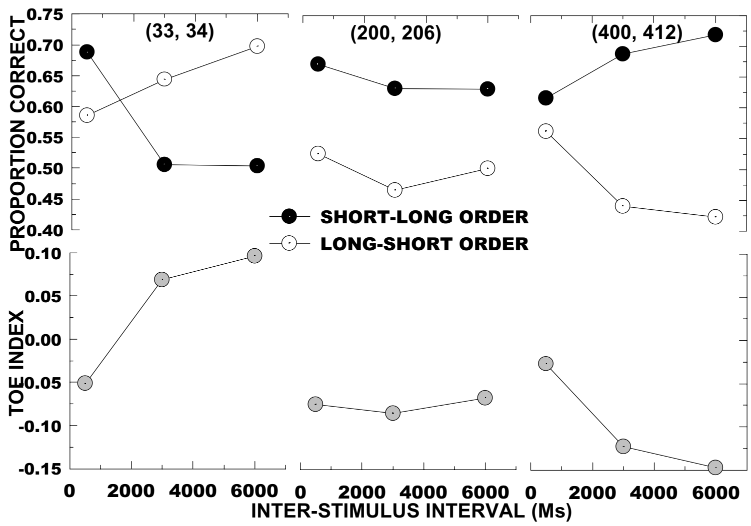


Figure 2. A three-way interaction between ISI, presentation order, and stimulus pair in terms of proportion correct. A two-way interaction between ISI and stimulus pair in terms of Jamieson and Petrusic's (1975) T_e index of TOE.

Differential sensation weighting analysis. The model successfully fit data at the 500 ms level of ISI, $F(2,3) = 24.64$, $p < .05$, $R = .97$, $R_{press}^2 = .71$. It also fit data at t[RM8]he 3000 ms level [$F(2,3) = 116.40$, $p < .01$, $R = .99$, $R_{press}^2 = .94$], and at the 6000 ms level, $F(2,3) = 92.50$, $p < .01$, $R = .99$, $R_{press}^2 = .91$. All coefficients were once again significant to the p =.01 level, except for a_0 at the 500 ms ISI. The RL was calculated to be 178 pixels, which is higher than the geometric mean of 140 pixels. Weight ratios (s_1/s_2) for the 500 ms, 3000 ms, and 6000 ms ISIs were 1.00, 0.964, and 0.965, respectively.

Wave theory analysis. Once again, Equation 2 was fit to the data. The regression performed at the 500 ms ISI level was significant, $F(1,4) = 30.30$, $p < .01$, $R = .94$, $R_{press}^2 = .78$. However, the model did not significantly fit the 3000 ms level, $F(1,4) = 2.85$, $p = .17$, $R = .65$, $R_{press}^2 < 0$, or the 6000 ms level, $F(1,4) = 2.31$, $p = .20$, $R = .61$, $R_{press}^2 < 0$.

Discussion

If one is to attribute TOE effects to either perceptual factors [RM9]or decisional factors, then the current experiments favour the perceptual locus. In both experiments, an interaction between stimulus level, presentation order and ISI was observed. TOEs for short and long line pairs moved in opposite directions as ISI lengthened. Other researchers studying the comparison of visual length have obtained similar results (e.g., Inomata, 1959). These results can only be explained in terms of decisional bias if two conditions are met: i) bias shifts in the absence of feedback, and ii) bias shifts systematically over time.

Hellström's differential sensation weighting model fit the proportion correct data well at all ISIs. The RL, an internal anchor affecting judgment, was greater than the geometric series mean in both experiments. In the first experiment, it was also greater than the arithmetic series mean. Unfortunately, there is no ready explanation for why similar stimulus sets should produce greatly divergent RLs. It appears that subjective scale values assimilated towards the RLs, and that RLs influenced first stimuli to a greater extent than second stimuli. The systematic shift in coefficients of ISI suggests that the model's weighting parameters co[RM10]vary with sensitivity.

Link's wave theory model provided a good fit to the data obtained for the shortest level of ISI. However, the wave theory model did not fit the data associated with the longer ISIs. It is possible that this failure was due to inappropriate model selection. Ideally, one should evaluate the wave theory model presented in Equation 3. This model has a sensitivity parameter ($\Delta S/S$) that is likely to vary over ISI. However, if one uses a sensitivity parameter rather than a stimulus-based variable, then there is no *a priori* variable, and no way to make quantitative predictions prior to data collection.

There are obvious avenues of improvement for both the differential sensation weighting model[RM11] and the wave theory model. The differential sensation weighting theory would have greater face validity if factors underlying the weighting parameters and RLs were revealed. Wave theory is largely dependent upon models that are solely comprised of parameters, and is in need of models that feature *a priori* variables. Both theories should eventually be expanded to explain contextual effects.

References

Ellis, H. D. (1973). An examination of the adaptation-level theory account of time errors. *Journal of Experimental Psychology*, 98(1), 160-163.

Fechner, G. (1966). *Elements of psychophysics*. (Transl. By H. E. Adler). New York: Holt, Rinehart and Winston.

Geertsma, R. H. (1958). Time-order errors in comparative judgments of hurtfulness. *Journal of Experimental Psychology*, 55(3), 284-288.

Hellström, Å. (1977). Time errors are perceptual. An experimental investigation of duration and a quantitative successive-comparison model. *Psychological Research*, 39, 345-388.

Hellström, Å. (1985). The time-order error and its relatives: mirrors of cognitive processes in comparing. *Psychological Bulletin*, 97(1), 35-61.

Hellström, Å. (2000). Sensation weighting in comparison and discrimination of heaviness. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 6-17.

Inomata, S. (1959). The analysis of conditions behind the time-error in the successive comparison of visual length. *Japanese Psychological Research*, 7, 36-53.

Jamieson, D. G. & Petrusic, W. M. (1975). Presentation order effects in duration discrimination. *Perception & Psychophysics*, 17, 192-207.

Link, S. W. (1992). *The Wave Theory of Difference and Similarity*. Hillsdale: Lawrence Erlbaum associates.

Needham, J. G. (1934). The time-error in comparison judgments. *Psychological Review*, 31(4), 229-243.

Petrusic, W. M., Harrison, D. H. & Baranski, J. V. (2004). Long-term memory for elementary visual percepts: memory psychophysics of context and acquisition effects. *Perception & Psychophysics*, 66(3), 430-445.

Rambo, W. W. & Johnson, E. L. (1964). Practice-effects and the estimation of adaptation[RM12]-level. *American Journal of Psychology*, 77(1), 106-110.

Acknowledgment

The work was supported by the Natural Sciences and Engineering Research Council of Canada through a grant made to William M. Petrusic.

MECHANORECEPTIVE MECHANISMS TO DETERMINE THE SHAPE OF THE DETECTION-THRESHOLD CURVE PRESENTING TANGENTIAL VIBRATIONS ON HUMAN GLABROUS SKIN

Tetsu Miyaoka
Shizuoka Institute of Science and Technology, Japan
miyaoka@cs.sist.ac.jp

ABSTRACT

The purpose of the present study was to measure vibrotactile thresholds presenting normal and tangential vibrations to the skin and to investigate the mechanoreceptive mechanisms that determine the tangential-threshold-curve shape. Two experiments were performed, in which the normal- and tangential-vibrotactile thresholds, respectively, were measured for seven subjects. When the stimulus frequencies were between 100 and 350 Hz, the normal- and tangential-threshold curves were U-shaped, and were similar to each other. Below 50 Hz the tangential-threshold curve was steeper than the normal-threshold curve. The results of these experiments, as well as those reported by Miyaoka (2004), indicate that the slowly adapting type II unit determined the shape of the tangential-threshold curve when the stimulus frequencies were below 100 Hz.

Many studies have been conducted to measure detection thresholds presenting vibrations on the skin (Gescheider, 1976; Miyaoka et al., 1985; Mountcastle et al., 1972; Verrillo, 1963, 1966, 1968). Bolanowski et al. (1988) identified four independent tactile channels based on psychophysical and neurophysiological studies and subsequent studies confirmed their findings (Gescheider et al., 2002). These studies, however, used normal vibrations to the skin surface. Very few studies have involved tangential vibrations to the skin surface (Miyaoka, 2004; Miyaoka et al., 2001).

The purposes of the present study were to measure vibrotactile thresholds presenting normal and tangential vibrations to the skin surface with small-sized contactors and to investigate the mechanoreceptive mechanisms that produce the tangential thresholds.

EXPERIMENT 1

In Experiment 1, vibrotactile thresholds were measured at the distal pad of the index finger presenting normal-sinusoidal vibrations to the skin surface. The purposes of Experiment 1 were to measure detection thresholds for the normal vibrations with a small contactor and to compare the obtained thresholds with the normal thresholds reported in previous studies. Experiment 1 and Experiment 2 were performed concurrently for the same subjects in order to avoid learning effects in the experiments.

Method

Subjects: Seven subjects, six males and one female in their twenties, participated in the experiment.

Stimuli and Apparatus: Normal-sinusoidal vibrations to the skin surface were adopted as stimuli. The frequencies of the stimuli were 4, 8, 16, 32, 50, 75, 100, 200, 250, and 350 Hz. The presenting time of each stimulus was 1000 ms and the rise and fall times were 44 ms, respectively. The normal stimuli were presented to the glabrous skin with a contactor of 2.5 mm in diameter. The contactor was attached to a vibrator (EMIC, 512A) that was assembled to produce normal vibrations. The contactor was set at the center of a 4.5 mm hole in a rigid surround. The distance between the edge of the contactor and the rigid surround was 1 mm, and the contactor was adjusted to the same height as the surface of the rigid surround. The temperature of the index finger, the contactor, and the rigid surround was maintained at 35ºC using a temperature-control device.

Procedure: The subject was seated and placed his/her left hand on the contactor and rigid surround. Normal vibrations were presented at the distal pad of the left index finger of the subject. The subject responded by pressing a switch on a response box according to the two-alternative-forced-choice technique. Stimulus amplitudes were changed by the procedure of parameter estimation by sequential testing (PEST; Taylor et al., 1967). For each subject, four measurements were conducted for each stimulus frequency. Thus, the total number of experimental trials for each subject was 40.

Results and discussion

Average normal thresholds were calculated for the seven subjects. The average threshold curve is shown in Fig. 1. The open circles in the figure denote the average thresholds and the vertical bars for each symbol denote standard deviations. The threshold curve obtained in Experiment 1 decreased gently between 4 and 32 Hz, became flat between 32 and 100 Hz, and showed a U-shaped between 100 and 350 Hz. The shape of the curve was typical of curves obtained using small contactors (Verrillo, 1963, 1966, 1968). Taking the previous studies into consideration, it was concluded that the fast adapting type I unit (FA I) contributed to produce the shape to the threshold curve below 100 Hz, and the fast adapting type II unit (FA II) determined the shape above 100 Hz.

The normal-threshold curves obtained in Experiment1 and reported by Miyaoka (2004) are shown in Fig. 2. The open circles denote the normal thresholds of Experiment 1, which were measured using the contactor having the diameter of 2.5 mm, and the open diamonds denote those reported by Miyaoka (2004), which were measured using the contactor having the diameter of 8 mm. Contactor size is known to strongly affect the shape of normal-threshold curves. When the stimulus frequencies were between 4 and 32 Hz, the 2.5-mm-contactor thresholds were smaller than the thresholds measured using the 8-mm contactor. However the two threshold curves crossed at approximately 50 Hz, and above 75 Hz the 8-mm-contactor thresholds became lower than the 2.5-mm-contactor thresholds. Below 32 Hz, the FA I is believed to be the mechanoreceptor that is responsible for the shape of threshold curves. The FA I is very sensitive to the gap width between the contactor and the surround. The gap width in Experiment 1 was 1 mm, and that reported by Miyaoka (2004) was 2 mm. The difference in gap width was thought to give the different thresholds. Over 100 Hz, the FA II is believed to determine the threshold values. The FA II shows clear spatial summations. Therefore, the 8-mm-contactor thresholds became smaller than the 2.5-mm-contactor thresholds.

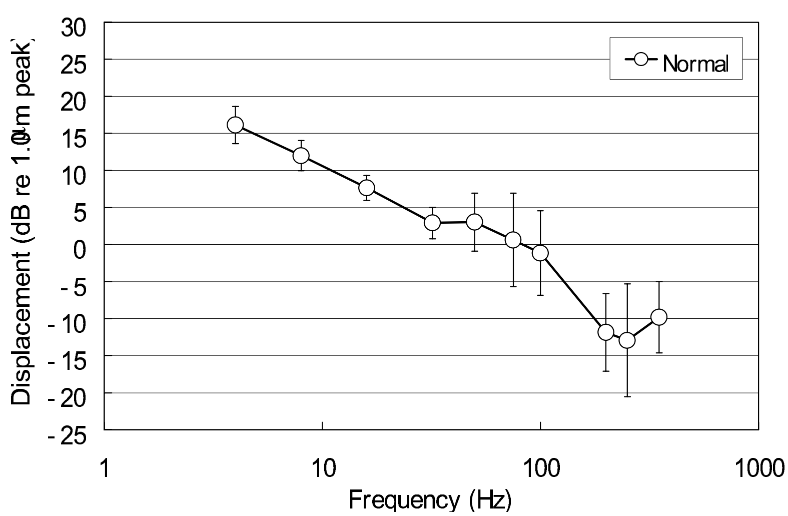


Fig. 1. Threshold curve measured at the distal pad of the index finger presenting normal vibrations to the skin. Vertical bars denote standard deviations.

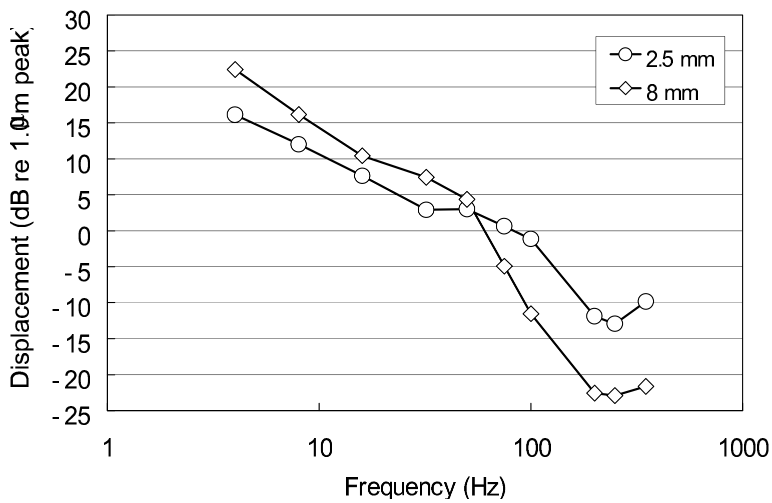


Fig. 2. Normal-threshold curves obtained in Experiment 1 (open circles) and reported by Miyaoka (2004) (open diamonds). The threshold curve obtained in Experiment 1 was measured using a 2.5-mm contactor, whereas that reported by Miyaoka was measured using an 8-mm contactor.

EXPERIMENT 2

In Experiment 2, vibrotactile thresholds were measured at the distal pad of the left index finger presenting tangential-sinusoidal vibrations with a small contactor to the skin surface. The purposes of Experiment 2 were to measure tangential thresholds and to compare the obtained thresholds with the tangential thresholds reported by Miyaoka (2004). As mentioned above, Experiment 2 was executed concurrently with Experiment 1 for the same subjects.

Method

Subjects: Seven subjects took part in Experiment 2. All subjects participating in this experiment also participated in Experiment 1.

Stimuli and Apparatus: Tangential-sinusoidal vibrations to the skin surface were adopted as stimuli. The frequencies, presenting time, rise-fall times of stimuli, and the devices used for controlling the amplitudes of stimuli were identical to those in Experiment 1. The tangential stimuli were presented with a contactor having a diameter of 2.5 mm. The contactor was attached to a vibrator (AKASHI, MEE-025) that was assembled to produce tangential vibrations. The contactor was set at the center of a 4.5-mm hole in a rigid surround. The distance between the edge of the contactor and the rigid surround was 1 mm, and the contactor was adjusted to the same height as the surface of the rigid surround, as in Experiment 1. The temperature of the index finger, the contactor, and the rigid surround was maintained at 35ºC using the same type of device used in Experiment 1.

Procedure: The subject was seated and placed his/her left hand on the contactor and rigid surround. Tangential vibrations were presented at the distal pad of the left index finger of the subject. The subject responded by pressing a switch on the response box. The response box and the response procedure were the same as in Experiment 1. Amplitudes of stimuli were changed by the PEST procedure. For each subject, four measurements were performed for each stimulus frequency. Therefore, the total number of experimental trials for each subject was 40.

Results and discussion

Average tangential thresholds were calculated for seven subjects. The average tangential-threshold curve is

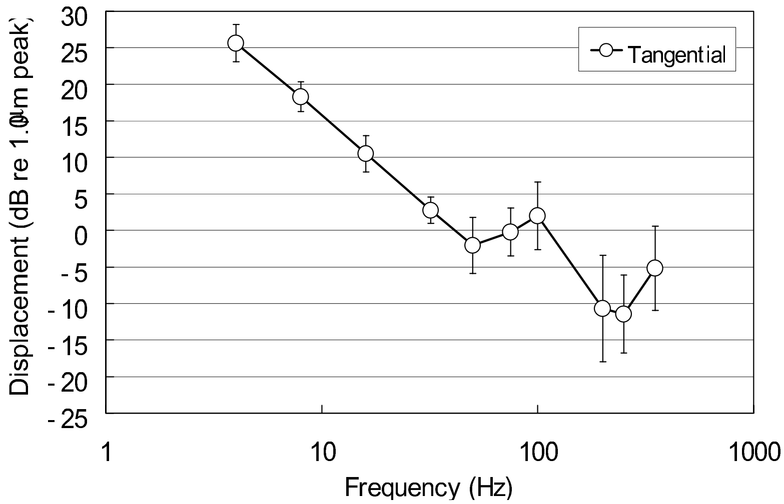


Fig. 3. Threshold curve measured at the distal pad of the index finger presenting tangential vibrations to the skin. Vertical bars denote standard deviations.

shown in Fig. 3. The open circles in the figure denote the average thresholds, and the vertical bars for each symbol denote standard deviations. The threshold curve decreased linearly from 4 to 50 Hz, increased gently until 100 Hz, and decreased again above 100 Hz. A U-shaped curve was observed between 100 and 350 Hz. The shape of the tangential-threshold curve indicates that at least two types of mechanoreceptors determine the shape of the curve. Based on previous studies, the FA II is believed to be the mechanoreceptor above 100 Hz (Bolanowski et al., 1988; Gescheider et al., 2002; Greenspan et al., 1996; Miyaoka, 2004; Verrillo, 1963, 1966).

Below 100 Hz, determining the participating mechanoreceptor becomes difficult. The exponent of the power function, fitted to the thresholds between 4 and 50 Hz, was –1.22. Exponents of power functions become

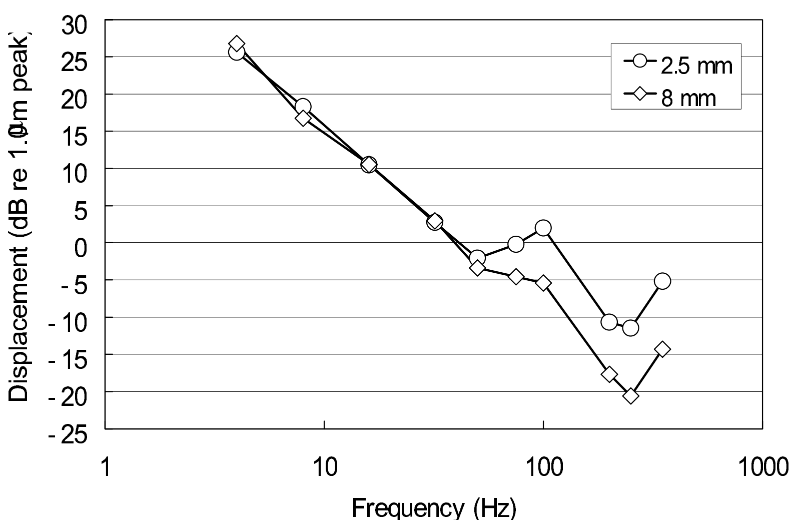


Fig. 4. Tangential threshold curves obtained in Experiment 2 (open circles) and reported by Miyaoka (2004) (open diamonds). The threshold curve in Experiment 2 was measured using a 2.5-mm contactor, whereas that reported by Miyaoka was measured using an 8-mm contactor.

approximately –1 and –2 when the participating mechanoreceptors are the FA I and FA II, respectively. The exponent of the present experiment differs from both values.

The tangential-threshold curves obtained in Experiment 2, together with those reported by Miyaoka (2004), are shown in Fig. 4. Open circles and open diamonds denote the tangential thresholds obtained in Experiment 2 and reported by Miyaoka (2004), respectively. The diameters of the contactors were 2.5 mm and 8 mm for Experiment 2 and Miyaoka (2004), respectively. The two threshold curves overlapped between 4 and 50 Hz, and were U-shaped but not overlapped between 100 and 350 Hz. The overlapping parts indicate that the same type of mechanoreceptor takes part in producing the curve slopes. Miyaoka (2004) suggested that the slowly adapting type II unit (SA II) contributed to determine the shapes of tangential-threshold curves. The slope of the curve in Experiment 2 was –1.22 and the overlap of the two curves showed no gap effect because the gap between the contactor and the surround was 1 mm in Experiment 2 and was 2 mm in Miyaoka (2004). These two results indicate that the participating mechanoreceptor is not the FA I.

DISCUSSION OF EXPERIMENTS 1 AND 2

The normal- and tangential-threshold curves measured in Experiments 1 and 2, respectively, for the same subjects are shown in Fig. 5. Open squares and open circles in the figure denote normal and tangential thresholds, respectively.

The normal- and tangential-threshold curves were similar when the stimulus frequencies were between 100 and 350 Hz. They were U-shaped and the participating mechanoreceptor was believed to be the FA II. When the stimulus frequencies were below 100 Hz, the slopes of normal and tangential curves became different from each other. The exponents of the power functions fitted to the two curves were –0.74 and –1.22 for the normal and tangential curves, respectively. The normal and tangential thresholds showed statistically significant differences at 4, 8, and 50 Hz (4 Hz, $t = –5.15$, $p < 0.01$; 8 Hz, $t = –4.46$, $p < 0.01$; 50 Hz, $t = 2.77$, $p < 0.05$). The tangential slope, –1.22, was similar to the slope of the SA II reported by Bolanowski et al. (1988). Furthermore, the tangential thresholds had no reaction to the gap change between the contactor and the surround. These findings do not support the hypothesis that the mechanoreceptor that determines the shape of the tangential thresholds is the FA I. The results of Experiments 1 and 2 and those reported by Miyaoka (2004) suggest that the SA II is the mechanoreceptor that takes part in producing the shape of tangential-threshold curve below 100 Hz.

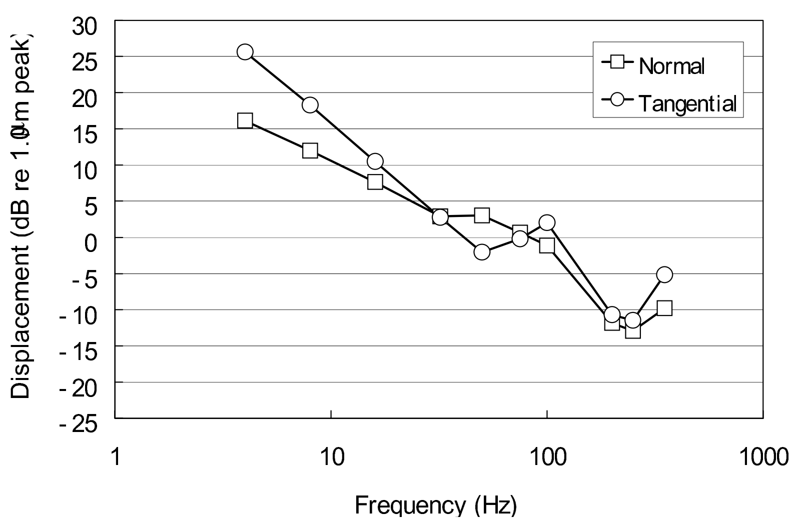


Fig. 5. Threshold curves measured in Experiment 1 (open squares) and Experiment 2 (open circles). Open squares and open circles denote normal and tangential thresholds, respectively.

CONCLUSION

The normal- and tangential-threshold curves were measured at the distal pad of the index finger using the 2.5-mm contactor. The shapes of the curves were differed between 4 and 100 Hz, and were U-shaped and similar to each other between 100 and 350 Hz. The normal-threshold-curve shape is determined by the FA I and FA II. The tangential-threshold-curve shape appears to be determined by two mechanoreceptors. One is thought to be the FA II, which contributes to determine the curve shape between 100 and 350 Hz, and the other appears to be the SA II, which determines the curve shape between 4 and 100 Hz.

Further studies are needed in order to determine clearly what types of mechanoreceptors contribute to determine the shape of the tangential-threshold curve.

ACKNOWLEDGMENTS

This work was supported in part by Grants-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science, Nos.11410027 and 14310043, and by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science, No.14510109.

REFERENCES

Bolanowski, S. J., Gescheider, G. A., Verrillo, R. T., and Checkosky, C. M. (1988). Four channels mediate the mechanical aspects of touch. *Journal of the Acoustical Society of America*, **84**, 1680-1694.

Gescheider, G. A. (1976). Evidence in support of the duplex theory of mechanoreception. *Sensory Processes*, **1**, 68-76.

Gescheider, G. A, Bolanowski, S. J., Pope, V., and Verrillo, R. T. (2002). A four-channel analysis of the tactile sensitivity of the finger tip: frequency selectivity, spatial summation, and temporal summation. *Somatosensory and Motor Research*, **19**, 114-124.

Greenspan, J. D. and Bolanowski, S. J. (1996). The psychophysics of tactile perception and its peripheral physiological basis. In L. Kruger (Ed.), *Pain and Touch*, edited by, Academic Press, San Diego, pp.25-103.

Miyaoka, T., Mano, T., and Fukuda, H. (1985). Threshold curves and equal-sensation contours for low frequency vibrotactile stimuli on the human glabrous skin. *Japanese Psychological Research*, **27**, 145-153.

Miyaoka, T. and Takahashi, Y. (2001). Measurements of detection thresholds when presenting normal and tangential vibrations to the skin. *The Bulletin of the Shizuoka Institute of Science and Technology*, **11**, 175-185. (In Japanese with English abstract)

Miyaoka, T. (2004). Measurements of detection thresholds presenting normal and tangential vibrations on human glabrous skin. *Proceedings of the Twentieth Annual Meeting of the International Society for Psychophysics*, 465-470.

Mountcastle, V. B., LaMotte, R. H., and Carli, G. (1972). Detection thresholds for stimuli in humans and monkeys: comparison with threshold events in mechanoreceptive afferent nerve fibers innervating the monkey hand. *Journal of Neurophysiology*, **35**, 122-136.

Taylor, M. M., and Creelman, C. D. (1967). PEST: Efficient Estimates on Probability Functions. *Journal of the Acoustical Society of America*, **41**, 782-787.

Verrillo, R. T. (1963). Effect of contactor area on the vibrotactile threshold. *Journal of the Acoustical Society of America*, **35**, 1962-1966.

Verrillo, R. T. (1966). Effect of spatial parameters on the vibrotactile threshold. *Journal of Experimental Psychology*, **71**, 570-575.

Verrillo, R. T. (1968). A duplex mechanism of mechanoreception. In D. R. Kenshalo (Ed.), *The skin senses*. Spring field, Ill.: Charles C Thomas, pp.139-159.

RESPONSE CHOICE STRUCTURE AND DIMENSIONAL ASYMMETRY

John S. Monahan
Department of Psychology, Central Michigan University, Mt. Pleasant Michigan 48859 USA, monah1js@cmich.edu

Abstract

The effects of response choice structure on dimensional asymmetry in Stroop tasks were investigated. Color was relevant for one set of participants; words were relevant for another. Participants responded by touching labels that were color patches or words and were fixed or varied on each trial. Fixed color labels yielded no dimensional asymmetry in interference. Other labels did. The results were interpreted as showing the effect of response related, rather than stimulus, factors in determining dimensional asymmetry.

Garner and Felfoldy (1971) demonstrated a set of speeded classification tasks that yield patterns of dimensional interaction which can be used to discriminate between integral and separable combinations of stimulus dimensions. Integral dimensions yield a speed gain with redundant combinations and interference with variation of the irrelevant dimension, whereas separable dimensions do not. A third dimensional combination mentioned by Garner allows no interference with irrelevant variation, but does produce facilitation. A fourth potential combination that would cause interference with little or no facilitation was not considered. Integral stimuli have mandatory processes because the dimensions are strongly combined. Separable stimuli have optional processes because they there is no stimulus-based need to combine the information from the dimensions. Thus clearly non-integral stimuli, such as Stroop stimuli, have the mandatory, but not stimulus-based, process the produces interference with irrelevant variation and the optional process of little or no facilitation. Optional processing is not dependent on stimulus concepts, whereas mandatory processes are.

A number of experimenters (c.f. MacLeod, 1990) have shown that Stroop stimuli demonstrate interference in color naming if the color words are incongruent and, sometimes, a small speed gain if the colors and words are combined redundantly. When the task is word reading, rather than color naming, the interference and speed gain often vanish. In Garnerian terms, color naming with Stroop stimuli may be mandatory stimulus processing, but word naming is optional stimulus processing. These results fit none of Garner and Felfoldy's categories.

Common models of the Stroop effect assume separate, i.e. optional, dimensional processing. For example, Eriksen and Eriksen's (1974) response competition model assumes that colors and words are processed separately and that the responses associated with each aspect prime separate responses. Another model, automaticity (Logan, 1980), is based on the idea that word reading is automatic, separate from, and has primacy over color naming(LaBerge & Samuels, 1974). Sugg and McDonald's (1994) version of the translation model assumes that color and word processing occur in separate cognitive structures as does the processing necessary for responses appropriate the each stimulus dimension. Thus the typically found asymmetric interaction of dimensions is usually thought to be based on response, not stimulus, processes.

Durgin (2004), on the other hand, along with a few others, argues that Stroop interference is based on internal processes that precede response processes rather than response processes themselves. Melara and Algom (2003) presented a general model of dimensional processing in which response processes as well as other factors, including stimulus factors, affect dimensional asymmetry. Thus the suggested sources of variability in Stroop results is relatively broad, yet researchers most often argue that response-related processes affect the results of Stroop experiments.

Stroop experiments differ from Garner experiments in both the stimulus sets used to gather data and the resulting analyses. Garner experiments are typically based on sets of four stimuli that vary symmetrically in two dimensions. The four control stimulus sets are composed of pairs of stimuli that vary in a single dimension. The two orthogonal sets, one for each dimension, are composed of all four stimuli, for which the task is to identify the level of one dimension while ignoring the other. The two redundant sets are each composed of two stimuli that vary in both dimensions. Each of the eight stimulus sets are presented separately in random order. From the results, two measures are derived: filtering (or selective attention) interference (orthogonal RT minus control RT) and redundancy gain (control RT minus redundant RT). Stroop stimuli, on the other hand, have between two and six levels for each dimension, although the typical number is four. There are only three types of Stroop stimuli: neutral, one dimension varies whereas the other assumes a value that is not one of the response choices; incongruent, color and word do not agree; and congruent-color and word agree. These sets do not correspond exactly to any of the Garner sets, although there are similarities. Stimulus sets may be presented separately but more often ate presented all together. Three measures are derived: interference (incongruent RT minus neutral RT), facilitation (neutral RT minus congruent RT), and congruity (incongruent RT minus congruent RT). Thus the Stroop measures, while somewhat similar to some Garner measures, are not identical to them. Stroop interference and Garner interference are not the same; Stroop's is probably inflated in comparison to Garner's because Garner's orthogonal set includes both incongruent and congruent stimuli. Although Stroop facilitation and Garner redundancy gain have a common thread -- dimensional correlation which leads to greater discriminability -- they are measured differently. Nevertheless, some comparison of Stroop results to Garner constructs may be useful, particularly dimensional asymmetry.

Dimensional Asymmetry and Response Modality

Using vocal responding and presenting stimulus sets separately, Stroop (1935) found significant interference both of words interfering with color identification and of colors interfering with word identification. He did not report the later probably because color interference was only about 5% of word interference. Monahan (unpublished manuscript) replicated these results in experiments in which each stimulus type was tested separately and always in the same order. Both vocal and key press responding were tested. Key tops were colored. With vocal responding there was significant interference with either words or colors as the relevant dimension, but the level of interference was much greater with color as the relevant dimension. With key press responding there was also significant interference with either color or words as the relevant dimension, but that interference was essentially equal; there was no dimensional asymmetry in interference. Vocal responding was faster than key press responding except for color relevant incongruent stimuli. Color relevant interference was greater with vocal than key press responding. Word relevant interference was greater with key press responding. With vocal responding, color relevant facilitation and word relevant negative facilitation was found. With key press responding, word relevant facilitation was found. Negative facilitation is not predicted by any model.

With key press responding, color interference with word identification was as great as word interference with color identification, whereas with vocal responding, color interference was much less than word interference. Thus response modality has an effect on relative magnitude of dimensional interaction. Because changing the response modality from vocal to key press, rather than changing the stimuli, eliminated the asymmetry in interference, the effect is probably not related to stimulus processing per se. Instead it may be related to processing beyond initial stimulus processing. An obvious question is whether manual responding always eliminates interference asymmetry or are there other aspects to consider. The current experiment varies response parameters while keeping response modality constant to test the effect of response choice structure on dimensional asymmetry.

Table 1. RT and Facilitation and Interference with separate presentation of stimulus types.

	Relevant		Stimulus Type				
Experimenter	Dimension	n	Congruent	Neutral	Incongruent	Facilitation	Interference
Stroop Vocal	Color	100		633 (6.80)	1103 (11.8)		470* (19.3)
Monahan Vocal	Color	85	503 (17.5)	519 (14.5)	695 (23.3)	16* (13.5)	176* (19.6)
Monahan Key	Color	85	585 (20.8)	592 (20.4)	689 (23.7)	7 (16.5)	97* (26.6)
Stroop Vocal	Word	70		410 (3.66)	433 (4.65)		23* (8.37)
Monahan Vocal	Word	85	505 (17.2)	472 (15.3)	510 (16.0)	-33* (10.4)	38* (9.5)
Monahan Key	Word	85	595 (22.8)	622 (19.2)	724 (32.7)	27* (16.7)	102* (18.1)

Note: Stroop's data are converted to mean time per item. Times are in ms and 95% confidence interval widths are given in parentheses.
* Significant interference or facilitation, = 0.05.

One of the features of Garnerian experiments is the visibility of the response categories at all times. In the current experiments, participants responded by touching a labeled area of a touch screen monitor. Thus the participants could see the labels while responding. Two types of labels were tested, color patches and color words in black. Area labels either changed on each trial, assuring the need to look at the label while responding, or remained the same through out the experiment, allowing memorization of the label locations, which could reduce the need to look at the labels on some trials.

One might expect that labels that match the relevant dimension might be responded to relatively quickly and that labels that do not change might also be responded to more quickly than those that do.

Method

Participants.

For color relevant stimuli, there were 120 participants who were assigned to four groups of 30. For word relevant stimuli, there were 152 participants who were assigned to four groups of 38. Participants received course credit.

Apparatus.

The programs were written in E-Prime (Schneider, Eschman, & Zuccolotto, 2002). Colors were standard E-Prime colors: red, yellow, blue, and lime.

Procedure.

Participants were instructed to indicate the presentation color of color words: red, yellow, blue, and lime (green), by pressing the appropriate label on the touch screen. Congruent (word and color matched), neutral (XXXX presented in color for color relevant stimuli and color words presented in gray for word relevant stimuli), and incongruent (word and color did not match) combinations were tested. For each of the two relevant dimensions, half the participants saw color patch labels, half saw word labels. For half the participants the locations of particular labels changed on each trial. For the other half, the locations of particular labels remained fixed. Participants were instructed to respond as quickly as possible while maintaining accuracy; RT and accuracy were measured. There was a practice block of 16 trials and three blocks of 36 data collection trials. Stimulus types were randomized within trial blocks. Participants were debriefed.

Results and Discussion

Color Relevant Results

With color as the relevant dimension, interference was substantial for all response conditions but was about half as much for fixed color labels as for the others. There was significant facilitation only with fixed word labels. As expected, color labels yielded faster RT than word labels. Fixed labels yielded faster RT than variable labels. These results are shown in Table 2.

Table 2. RT and Facilitation and Interference with Combined presentation of stimulus types.

	Relevant		Stimulus Type				
Label	Dimension	n	Congruent	Neutral	Incongruent	Facilitation	Interference
Color Fixed	Color	30	523 (32.5)	515 (29.9)	644 (51.2)	-8 (15.0)	129* (40.6)
Color Variable	Color	30	632 (34.7)	622 (34.7)	847 (63.1)	-10 (12.1)	225* (39.7)
Word Fixed	Color	30	772 (56.8)	808 (77.4)	1056 (78.0)	36* (23.4)	248* (44.0)
Word Variable	Color	30	978 (55.6)	983 (50.0)	1268 (75.4)	5 (24.8)	225* (40.2)
Color Fixed	Word	38	629 (27.6)	681 (31.0)	804 (43.5)	52* (15.2)	123* (28.7)
Color Variable	Word	38	650 (29.6)	727 (32.3)	847 (41.2)	77* (16.3)	120* (28.9)
Word Fixed	Word	38	704 (30.6)	725 (31.9)	761 (36.2)	21* (11.4)	36* (15.6)
Word Variable	Word	38	798 (32.6)	802 (35.9)	819 (36.2)	4 (16.0)	17* (13.7)

Note: Times are in ms and 95% confidence interval widths are given in parentheses.
*Significant facilitation or interference, = 0.05.

With color as the relevant dimension, interference occurred with color labels as well as word labels, but the underlying RT were different. Because there was interference with all labels, the color and word stimuli must have been processed to a level that caused interference. On the other hand, because the word labels were responded to more slowly than color labels, word response priming seems an unlikely source of word interference. Notice that the interference with fixed color label responding is similar to the level of interference with key press responding. With fixed labels, remembering the label locations is useful in speeding responses. With key press responding, remembering key locations is strongly encouraged to avoid constant looking away from the screen. Thus, the faster responding with fixed labels as well as the similar interference level with key press responding indicates a role for memory in dimensional interaction.

If Stroop dimensions are processed separately, stimulus processing should not cause interference except that it might direct attention asymmetrically. Response selection or priming or the responses themselves would seem likely causes of the interference. Similarly, it is likely that early stimulus processing is not affected by response processing. With variable labels, word responses lag comparable color responses by more than 300 ms. Presumably the physical act of making the response does not differ significantly between color and word labels, and thus, any differences in RT probably occur during response selection or priming. Word response primes would seem to arrive too late to influence color responding, yet Stroop interference is similar with color or word variable labels. Therefore a likely location for interference is in response selection.

Word Relevant Results

With words as the relevant dimension, color labels, fixed and variable, yielded both facilitation and interference. Word labels yielded interference, and fixed, but not variable, word labels yielded facilitation. Color labels yielded faster RT than word labels for redundant and neutral stimuli, but not for incongruent stimuli. Fixed labels yielded faster RT than variable labels for all stimuli. These results are shown in Table 2. Thus again response choice structure seems to have had an effect on dimensional interaction: Labels representing the irrelevant dimension yielded facilitation and interference. Labels representing the relevant dimension yielded considerably less interference and facilitation only sometimes.

Dimensional Asymmetry Results

A comparison of dimensional asymmetry and its underlying components as a function of response choice structure is shown in Table 3. For fixed color, fixed word, and variable word labels, RT was faster when the label matched the relevant dimension than when it did not. For variable color labels, RT was faster for the matching relevant dimension only for neutral stimuli, not for congruent or incongruent stimuli. Facilitation was significant for color labels with word relevant stimuli and for fixed word labels with either dimension relevant, without dimensional asymmetry. There was no facilitation with variable word labels. Only fixed word labels showed both significant facilitation and a lack of facilitation asymmetry. Variable color and fixed and variable word labels yielded significantly less interference with words as the relevant dimension than with color as the relevant dimension. With fixed color labels, however, there was no significant difference in interference between color and words as the relevant dimension. This result echoes the finding shown in Table 1 that key press labels yield similar interference with either color or words as the relevant dimension.

Table 3. RT, Facilitation, and Interference Differences by Label and Relevant Dimension

	Congruent			Neutral			Incongruent		
	Relevant Dim.			Relevant Dim.			Relevant Dim.		
Label	Color	Word	Difference	Color	Word	Difference	Color	Word	Difference
Color Fixed	523	629	160* (41.7)	515	681	166* (43.0)	644	804	160* (65.6)
Color Variable	632	650	18 (44.6)	622	727	105* (46.8)	847	847	0 (71.3)
Word Fixed	772	704	-68* (59.8)	808	725	-83* (57.8)	1056	761	-295* (78.7)
Word Variable	978	798	-180* (60.3)	983	802	-181* (58.9)	1268	819	-449* (76.7)

	Facilitation			Interference		
	Relevant Dim.			Relevant Dim.		
Label	Color	Word	Difference	Color	Word	Difference
Color Fixed	-8	52*	-60* (21.3)	129*	123*	6 (49.7)
Color Variable	-10	77*	-87* (20.3)	225*	120*	105* (49.1)
Word Fixed	36*	21*	15 (26.1)	248*	36*	212* (46.7)
Word Variable	5	4	1 (29.5)	225*	17*	208* (42.5)

Note: Times are in ms and 95% confidence interval widths are given in parentheses.
*Significant change in RT, facilitation, or interference, = 0.05.

The current results may be instructive in determining the response structure prerequisites to prevent interference asymmetry. These would appear to be twofold. First the response parameters must favor color naming, rather than word reading, presumably because word reading is more automatic (LaBerge & Samuels, 1974) than color naming. Thus vocal responding and word label patched should produce interference asymmetry. Second, the response parameters must either require or favor storing stimulus-response relations in memory. Thus both fixed color patches and key press responding should not produce interference asymmetry, but variable color or word patches should because they do not allow memory of stimulus-response relations.

Facilitation asymmetry was eliminated with fixed word labels, but fixed word labels yielded interference asymmetry. Interference asymmetry was eliminated with fixed color labels, fixed color labels yielded facilitation asymmetry. It would appear that Stroop stimuli will not yield Garnerian integral results. Because Stroop dimensions are probably processed separately, Garnerian result should not be expected.

Garner's labels, were not only always visible, but they also indicated the level of the relevant stimulus. Using Garner's logic, instead of comparing facilitation and interference with the same label, one should compare color relevant results with fixed color labels to word relevant results with fixed word labels. That comparison shows both facilitation and interference asymmetry. Although Garner might attribute the pattern of results to stimulus processes, such as state or process differences in discriminability, it is also possible that stimulus-response relations are the cause

Assuming that Stroop dimensions are separable, the question remains, how does interference occur? With integral stimuli, one hypothesis is that the stimulus is processed as a whole before it can be separated into its component parts. With Stroop stimuli, that does not seem to be the case because interference can be greatly reduced with word relevant responding by using word labels on the response patches. Therefore, as argued by most Stroop theorists, Stroop stimuli are most probably separable. If interference does not occur in stimulus processing, it must occur in response selection and processing. The results of the current experiment show that for Stroop stimuli, some of the interference occurs in response selection.

The current research also shows the usefulness of Garner's use of converging operations. Dimensional symmetry in interference in Stroop tasks can be achieved using fixed color labels. Dimensional symmetry in facilitation in Stroop tasks can be achieved using fixed word labels. What cannot be achieved is dimensional symmetry in both interference and facilitation using labels appropriate to the relevant dimension.

Acknowledgements

The data for this experiment were gathered by Amber Thompson. The study was supported in part by a CMU Faculty Research and Creative Endeavors Grant.

References

Durgin, F.H. (2004). Translation and competition among internal representations in a reverse Stroop effect. *Perception & Psychophysics, 65*(3), 367-678.
Eriksen, B.A. & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16*(1), 143-149.
Garner, W.R. & Felfoldy, G.L. (1971). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*(3), 225-241.
LaBerge, D. & Samuels, S.J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293-323.
Logan, G.D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*(4), 523-553.
MacLeod, C.M. (1990). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203.
Melara, R.D. & Algom , D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review, 110*(3), 422-471.
Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools, Inc.
Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.
Sugg, M.J. & McDonald, J.E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 647-675.

PSYCHOPHYSICAL APPROACH TO AUDITORY SEARCH IN FREQUENCY AND INTENSITY DIMENSIONS

Shuji Mori
Faculty of Sciences, Tokyo Metropolitan University, smori@comp.metro-u.ac.jp

Abstract

This paper reports a series of psychophysical experiments on the listener's search performance for a target tone embedded in a rapid tonal sequence. The experiments have demonstrated that the listener's performance, measured by the detection threshold and *d'*, improve as the target tone is increasingly separated in frequency and/or intensity from non-target (distractor) tones in the sequence. Furthermore, high-intensity distractor tones have large detrimental effects on the search performance when the target-distractor frequency separation is small, whereas the detrimental effect diminishes when the frequency separation is large. These results are considered in context of mechanisms of auditory attention (Leek, Brown, & Dorman, 1991; Hübner, 1993).

In auditory search, a listener listens for a pre-specified target tone among non-target, distractor tones which are presented simultaneously or in temporal sequence with the target tone. The term 'search' has not been used in auditory research until recently (Asemi, Sugita, & Suzuki, 2003; Cusack & Carlyon, 2003), and few studies have conducted quantitative, psychophysical investigation into the effects of acoustical variables such as frequency and intensity on the listener's ability to search for the target (Leek, Brown, & Dorman, 1991; Mori & Wong, 2004).

In the present study, a series of psychophysical experiments were conducted to examine the effects of frequency and intensity on the listener's search performance, with the methodologies adopted from the study of informational masking where a target tone is inserted in successive presentations of distractor tones (Watson, 1987).

Experiments

Stimulus tones were generated by a sound generation system (Tucker and Davis Technologies, Experimenter II) controlled by a personal computer (IBM, PS/V model2410), which also controlled timing and data collection. All tones were presented monaurally via a headphone (STAX, SR-Lambda PRO) to the listener's left ear. Responses were made on a custom-made response box which was connected to the computer via the interface of the sound generation system and also used to present warning signals and feedback on its LED display.

Target and distractors were sinusoidal tones of 46 msec in duration, with 3-msec rise and fall times for each tone. Figure 1 illustrates the frequency settings of the target and the distractor tones in the experiments reported here (except for the Condition E of Experiment 4). The target was always 800 Hz, and a set of eight distractors were separated in 1-semitone steps, four of the eight distractors being above 800 Hz and the other four below 800 Hz. The frequency separation between the target and the distractors was defined by the difference in semitone between the target and the nearest distractors in the upper and the lower groups of 4 tones. The intensity levels of the target and the distractors were all set at 75 dB SPL in Experiments 1 and 2, which investigated effects of frequency separation, while the intensity levels were varied systematically in Experiments 3 and 4 (see below).

Those tones were used to create 8-tone sequential presentations with no silent interval separating successive tones. There were two types of stimulus sequences, target-absent and target-present.

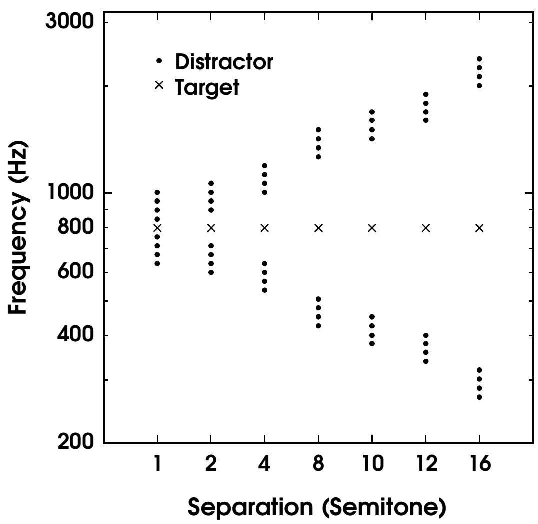


Fig. 1. Frequencies of target and distractor tones.

The target-absent sequences were composed of a set of eight distractors for a given frequency separation (Figure 1). The target-present sequences were composed of an 800-Hz target and seven distractors for a given frequency separation, by excluding either one of the two distractors closest to the target frequency (just above or below 800 Hz). For example, the two distractors closest to the target for the 2-semitone separation are 713 and 897 Hz, one of which was chosen randomly on each trial and not included in the target-present sequence. In the target-present as well as the target-absent sequences, 8 tones were ordered randomly in a sequence on every trial, with the constraint that the target and the nearest distractor were not placed at either the beginning or the end of the stimulus sequence (Leek et al., 1991; Watson, 1987).

In Experiments 1, 3 and 4, the listener's search performance was measured in terms of *d*'. These experiments were run in sessions separately for different conditions of frequency separations (Experiment 1), intensity separations (Experiment 3), and conjunctions of frequency and intensity separations (Experiment 4). In the sessions, listeners first listened to alternative presentations of a target tone alone and a target-present sequence for a given condition (the order of 8 tones were randomized every time) as often as they wished. They then ran 20-30 practice trials and 100 experimental trials, with a random half of the trials containing a target-present sequence and the other half a target-absent sequence. On each trial, a visual presentation of a 500-msec warning signal on the LED display of the response box preceded the presentation of a stimulus sequence. Given a one-interval forced-choice task, the listener was asked to detect the presence or absence of a target tone in the sequence and press a corresponding button on the response box. There was no time constraint for making a response, and following the response, accuracy feedback was provided on the LED display for 1000 msec. The inter-trial interval was 1000 msec. One session took 10 to 15 min.

d' was computed from proportions of hit and false alarms of 100 experimental trials in each session. In the present experiments, there were several sessions with 1.00 hit and/or 0.00 false-alarm proportions, which yield an infinite value of *d*'. To avoid this, we followed a convention suggested in the literature (e.g., Macmillan & Creelman, 1991) that 1.00 and 0.00 proportions are replaced respectively with $1 - 1/2N$ and $1/2N$, where N is the number of trials for the target presence, i.e. 0.99 and 0.01 in the present experiments. Therefore, the highest possible value of *d*' computed in the present study was 4.653.

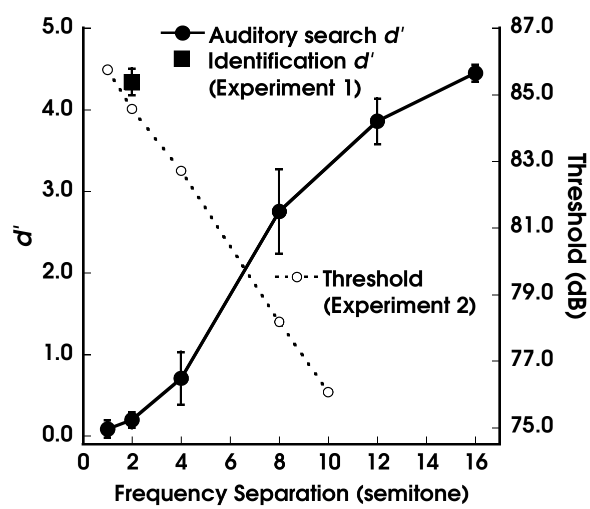
In Experiment 2, the listener's performance was measured in terms of the detection threshold using the QUEST staircase procedure (Watson & Pelli, 1983) which varied the intensity level of the target depending on the listener's response. We adopted the rule proposed by King-Smith, Grigsby, Vingrys, Benes, and Supowit (1994) to obtain the threshold corresponding to a detection probability of 63.7%. For a given frequency separation, the listener performed the task in a block of 100 trials (note that only the target-present sequences were presented); on the first trial, the target intensity was set at the level determined by the results of preliminary experiments using the method of limit, and the target intensity on the last trial was taken as the threshold level (for details, see Mori & Wong, 2004).

In the present experiments, there were a total of 7 participants, 6 graduate students (23-32 years old) and the first author (44 years old). Six of them took part in all the experiments reported here, and one student performed only in Experiments 1 and 2. The first author had participated in psychoacoustic experiments for more than 100 hours before those experiments, while none of the students had prior experience of psychoacoustic experiments and they were naïve to the purpose of the present study.

In Experiments 1 and 2, the listener's search performance was measured as a function of frequency separation between the target and the distractor tones. In Experiment 1, there was an additional measurement of the listener's identification performance for a set of 3 tones, the target (800 Hz) and its closest distractors of the 2-semitone separation (713 and 897 Hz), all of which were presented alone. In Experiment 3, the search performance was measured in terms of *d*' as a function of intensity separation, which was manipulated by increasing the target intensity with the distractor intensities fixed at 75 dB. Figure 2 summarizes the results of the three experiments. The results show clear effects of frequency and intensity separations on the search performance: the performance improves as the separations increase. In Experiment 1, the listeners' search performance was much worse than their identification performance, indicating that the difficulty of searching for the target among the distractors is not simply due to a limit on frequency identification or discrimination of the target and the distractors.

In Experiment 4, conjunctive effects of frequency and intensity separations were examined. The top panel of Figure 3 illustrates the intensity and frequency settings of five conditions of Experiment 4. The intensity separations were always 8 dB, i.e. 83 dB against 75 dB. The conditions A to D were employed for the 8-semitone frequency separation, while the conditions C to E were conducted for the 2- and 4-semitone separations. In the condition A, the 478-Hz and the 1340-Hz distractor tones, which were the second closest to the target frequency among the below-target and the above-target distractor groups respectively, were set at 83 dB while the other distractors and the target were 75 dB. In the condition B, the second and the fourth (most remote from the target) distractor tones (426 and 1503 Hz) were set at 83 dB while the other tones were 75 dB. The conditions C and D were similar to the conditions A and B, except that the target was also set at 83 dB. The condition E was employed for the 2- and 4-semitone separation. In this condition, the most remote distractor frequencies from the target were shifted to 478 and 1340 Hz and set at 83 dB, together with the target, while the other distractors were 75 dB. Those five conditions resulted in ten different intensity-frequency separations, which were tested in separate sessions, with their order counterbalanced across 6 participants.

The bottom panel of Figure 3 shows the results, the mean *d*'s of 6 participants for the five conditions. Also plotted are the results of the same participants in Experiment 1 and those for the 8-dB intensity separation in Experiment 2 that indicate their performance for the 0-dB and the 8-dB target-



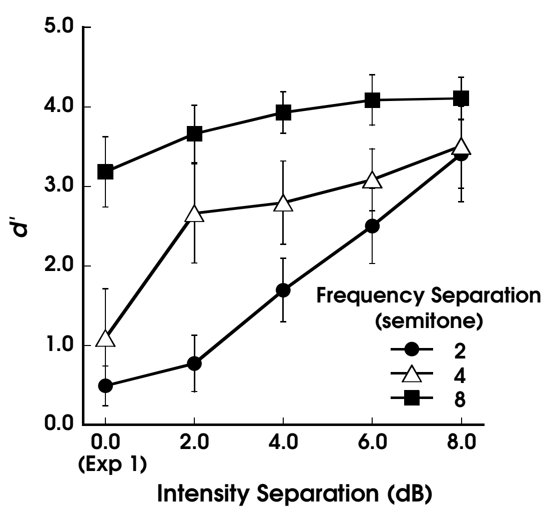


Fig. 2. Results of Experiments 1 and 2 (left panel) and of Experiment 3 (right panel).

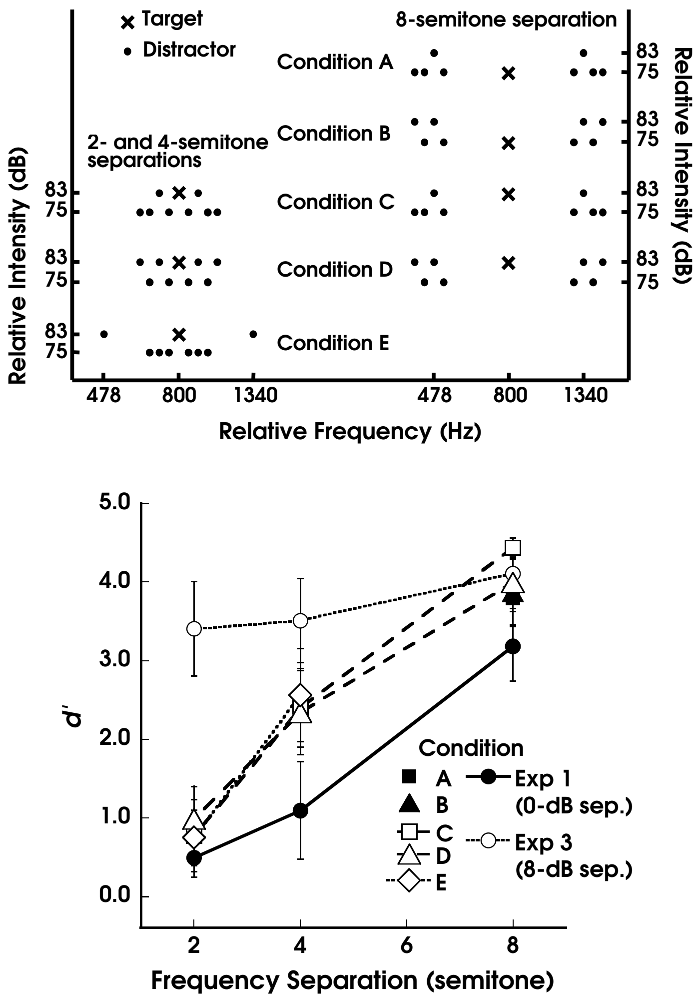


Fig. 3. Intensity and frequency settings (top panel) and results (bottom panel) of Experiments 4. Some results of Experiments 1 and 3 are plotted together in the bottom panel.

distractor separations, respectively, with no separation among the distractors. As can be readily seen in the figure, the listeners' performance for the five conditions of Experiment 4 were mostly determined by the frequency separation between the target and the distractors, with little effect of the number of high-intensity (83 dB) distractors (2 or 4) or target intensity level (75 or 83 dB). It is also evident that detrimental effects of intensity separations among distractors, as revealed by the comparison of the data of this experiment with those of Experiments 1 and 2, were dependent on the frequency separation. At the 8-semitone separation, *d*'s of the four conditions of this experiment were clattered together with those of Experiments 1 and 2, showing no indication of detrimental effects. At the 4-semitone separation, *d*'s of this experiment were in between those of Experiments 1 and 3. At the 2-semitone separation, *d*'s of this experiment were much lower than *d*' of Experiment 3 and comparable to that of Experiment 1, which indicates the lowest performance level with respect to target-distractor intensity separation.

Discussion

That the listener's search performance improved as the target was increasingly separated from the distractors in frequency (Experiments 1 and 2) or in intensity (Experiment 3) is consistent with what is called *release from informational masking* (Leek et al., 1991). Detrimental effects of informational masking on detection and discrimination performance diminish as the difference between the signal and the masker tones increases along some acoustical dimension (Leek & Watson, 1984; Spiegel & Watson, 1981). It has been suggested that the release from informational masking, as well as the detrimental effects of the masking, results from the operation of the listener's attention (Leek, 1986). Under informational masking, the listener's attention is distracted and not properly focused on the target. Separating the signal from the maskers along some acoustical dimension directs the listener's attention to a spectrotemporal location of the target. To quantify the operation of attention under informational masking, Leek et al. (1991) estimated an attentional filter, by fitting a roex function (Patterson & Moore, 1986; Patterson, Nimmo-Smith, Weber, & Milroy, 1982) to discrimination thresholds obtained as a function of frequency separation between the signal and the maskers in a way quite similar to the manipulation used in Experiment 2 of the present study. They demonstrated that the estimated attentional filter resembled a critical-band filter which is generally considered as reflecting peripheral processing of the auditory system (e.g., cochlea).

The effects of frequency and intensity separations on auditory search performance may also be modeled as the operation of the listener's attention, but in a different way from that proposed by Leek et al. (1991). Following Leek et al. (1991), we fitted a roex function to the detection thresholds obtained in Experiment 2 and found that the estimated filter was five times wider, in terms of equivalent rectangular bandwidth, than that reported in Leek et al. (1991) as well as a critical-band filter (for details, see Mori & Wong, 2004). Furthermore, some results of Experiment 4 are not compatible with the peripheral origin of attentional filter. In the condition C, the high-intensity distractors at 478 and 1340 Hz (see Figure 3) had little effect on search performance, whereas in the condition E, the same two distractors had clear detrimental effects, the effects being larger for the 2-semitone separation than for the 4-semitone separation. In other words, the high-intensity distractors at the same frequency locations had differential effects on search performance depending on the frequency separation of the stimulus sequence. These results are not readily explainable by single fixed-width filters such as critical-band filters.

It may be more appropriate to consider more flexible mechanisms of auditory attention: the width of the filter, and the number of filters the listeners monitor simultaneously, are variable under control of central processing of the auditory system (for a review, see Hübner, 1993). For example, let us assume that the filter width increases with the increasing degree of the listener's uncertainty about the spectrotemporal location of a target, and the listener's uncertainty is dependent on the frequency separation, not on the inclusion of high-intensity distractors, in the stimulus sequence. For a small separation, the listener's uncertainty is large, widening the filter to encompass the high-intensity distractors at distant frequencies. As a result, those distractors increase the overall output of the filter and make it difficult to discriminate between the stimulus sequences with and without a target. For a large separation, the listener's uncertainty is small, and the filter width does not encompass the high-intensity distractors, which are not effective on the output of the filter. Other attentional mechanisms may also account for the present and other findings of auditory search, and detailed quantitative analysis of those mechanisms are necessary in future research.

Acknowledgements

The research reported in this paper was supported by Grants-in-Aid for Scientific Research from the Japan Society for the Promotion of Science to S. M. (Grant No. 10239600) and to Yoshitaka Nakajima (Grant No. 14101001).

References

Asemi, N., Sugita, Y., and Suzuki, Y. (2003). Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal-horizontal plane. *Acta Acustica*, **89**, 346-354.

Cusack, R., and Carlyon, R. P. (2003). Perceptual asymmetries in audition. *Journal of Experimental Psychology: Human Perception and Performance*, **29**, 713-725.

Hübner, R. (1993). On possible models of attention in signal detection. *Journal of*

Mathematical Psychology, , 266-281.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., and Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: Theory, simulations, experimental evaluation and practical implementation. *Vision Research*, **34**, 885-912.

Leek, M. R. (1986). Directed attention in complex sound perception. In W. A. Yost & C. S. Watson (eds.), *Auditory processing of complex sounds* (278-288). Hillsdale, NJ: Erlbaum.

Leek, M. R., Brown, M. E., and Dorman, M. F. (1991). Informational masking and auditory attention. *Perception & Psychophysics*, **50**, 205-214.

Leek, M. R., and Watson, C. S. (1984). Learning to detect auditory pattern components. *Journal of the Acoustical Society of America*, **76**, 1037-1044.

Macmillan, N. A., and Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, NY: Cambridge University Press.

Mori, S., and Wong, W. (2004). Quantifying frequency dependence of auditory search. *Journal of Physiological Anthropology and Applied Human Science*, **23**, 303-306.

Patterson, R. D., and Moore, B. C. J. (1986). Auditory filters and excitation patterns as representations of frequency resolution. In B. C. J. Moore (ed.), *Frequency selectivity in hearing* (123-177). New York: Academic Press.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *Journal of the Acoustical Society of America*, **72**, 1788-1803.

Spiegel, M. F., and Watson, C. S. (1981). Frequency discrimination with components of well-learned patterns. *Journal of the Acoustical Society of America*, **69**, 223-230.

Watson, C. S. (1987). Uncertainty, informational masking, and the capacity of immediate auditory memory. In W. A. Yost & C. S. Watson (eds.), *Auditory processing of complex sounds* (267-277). Hillsdale, NJ: Erlbaum.

Watson, A. B., and Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, **33**, 113-120.

PSYCHOPHYSICAL THINKING IN BUSINESS: PRODUCTS & CONCEPTS

Howard R. Moskowitz
Moskowitz Jacobs Inc.
mjihrm@sprynet.com

Abstract

Psychophysical thinking informs both concept development and product design. With products the researcher systematically varies ingredients/processes and measures consumer reactions such as acceptance and sensory impressions. The product data generates a model showing how the formulations drive ratings. With concepts (and package designs) the researcher systematically varies the presence/absence of elements/pictures and measures consumer reactions such as acceptance or appropriateness. The concept data generate a model showing how the elements drive reactions. Concept and product research search for relations among variables, at once to understand and also to engineer consumer-acceptable products and messaging.

Psychophysical scaling and product modeling

About 40 years ago, as psychophysics was busy uncovering equations relating physical stimuli to perceived magnitude, sensory researchers in consumer product industries were watching and absorbing those efforts. The outcome, some ten years later in the mid 1970's was the development of psychophysical equations, not so much for model systems – the darlings of psychophysicists – but rather for real food items, such as pasta sauces, colas, coffees, cheeses, and the like. Psychophysical thinking informed applied sensory analysis, and as researchers moved on from discrimination testing to uncovering relations among variables, so did the corporate research.

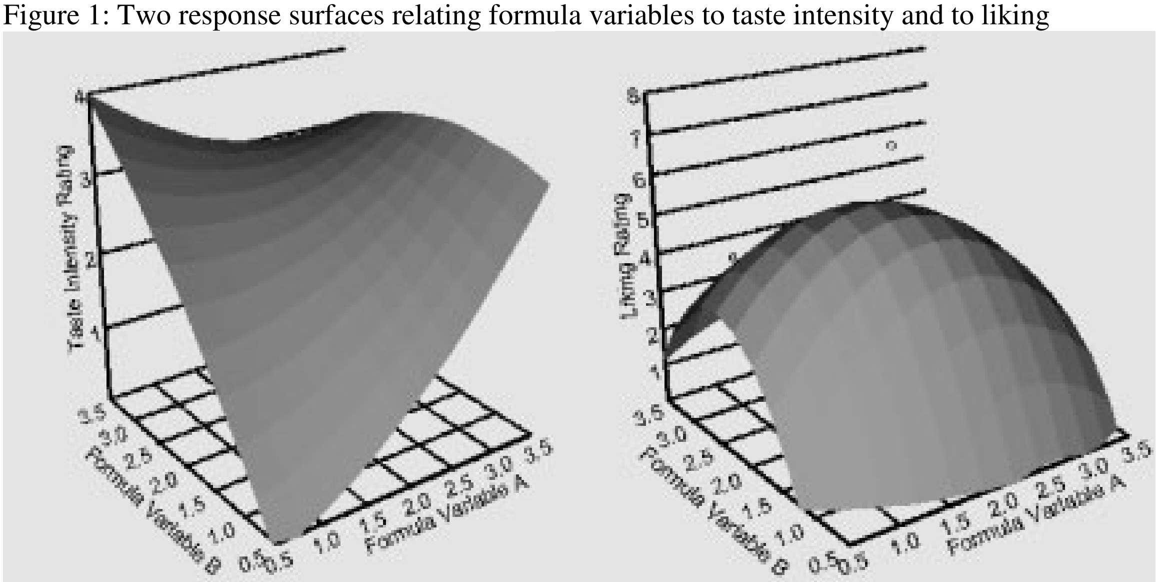
In some ways, however, applied research trumped psychophysics. Model systems can lead only so far in understanding how we respond to physical stimuli; food products are much richer. In the chemical senses, for instance, sensory magnitude and sensory hedonics are important. In foods, the measurable reactions go beyond magnitude and hedonics, onto image and appropriateness dimensions. The psychophysical richness is far greater with, say, a systematically varied cookie product because the product comes fully loaded with cognitive as well as sensory and hedonic baggage. The notion of 'a real European cookie' is neither a sensory nor a hedonic word, but allows psychophysical modeling for cookies. All the researcher need do is vary the ingredients systematically and instruct the observer to rate 'true European *cookieness*'. It's a surprisingly easy task; even if observers can't articulate their criteria, nonetheless they seem to do the task rapidly, effortless, and surprisingly reliably.

Psychophysicists look for 'laws' or regularities signaling the nature of our sensory/ cognitive systems. The equations spun out have meaning, in what Smitty Stevens used to call sensory, cognitive, or psychophysical mythology. The parameters were to be fussed over, interpreted, and of course argued. Applied sensory researchers were guided by the notion of regularities, not laws. It's hard to get worked up over a law of pasta sauce, another law of carbonated beverage, a third law of cookies, etc. To applied researchers the equations were just simple summaries of relations among variables. It's also a bit presumptuous. Yet, over time and with lots of experiments, the parameters became increasingly familiar; those parameters describing how sweetness changed with sweetener, how perceived bubbliness changed with carbonation. Applied sensory work, taking the psychophysics model as a useful metaphor for research design, in fact ended up with rules of thumb based upon the empirical models.

Applied sensory work went beyond today's psychophysics in two areas; interactions and reverse engineering. Psychophysicists work with mixtures, and have for more than eighty years. Mixtures of stimuli bring into play complexities of nature. Psychophysicists, especially taste and smell psychophysicists, have dealt with these mixtures, often following a slow, deliberate pace, to 'get the laws or regularities' right. There was time enough in one's career to invest in careful mixture research, varying all sorts of pairs of stimuli, measuring reactions, looking for enhancement or suppression. Working with pasta sauces, coffees, cookies, artificially sweetened beverages and the like plunges the applied sensory researcher into mixtures from day one. Without even a simple science that give rules, the applied researcher is supposed to create models that show how mixtures behave when they drive sensory perceptions of a product, such as a cookie (e.g., perceived sweetness, perceived hardness), and at the same time how they behave when they drive overall liking, authentic European flavor, etc. To cope with this demand the applied sensory researcher abandoned the rigorous search for interactions and the proper 'function form', settling down instead to work with simple quadratic polynomial equations of the form: Rating= $k_0 + k_1(A) + k_2(A^2) + k_3(B) + k_4(B^2) + k_5(AxB)$. More terms can be accommodated in functions appropriate for three, four, five, or even ten or more independent variables.

A sense of psychophysical thinking applied to product research can be obtained from Figure 1, which shows how two variables interact with each other to generate sensory ratings of taste/flavor intensity and liking, respectively. The researcher creates the combinations, gives them to the observers, with instructions to rate the combinations (sauces) on a set of attributes. After averaging the ratings from 10-20 observers, the researcher can create these curves because the noise cancels out. The psychophysicist would stop at this point, having shown how the ratings capture the ingredients. The interesting things here are that the curves look different because the attributes differ (although the ingredients are the same), and that one can use these data to identify optimal formulations. In a sense, validity is achieved by creating the actual formulation and testing it in the market.

Figure 1: Two response surfaces relating formula variables to taste intensity and to liking



Typically psychophysical thinking stops at establishing laws of nature. As S.S. Stevens was wont to say, the real job of psychophysics is to establish the stimulus. Afterwards it is simple to establish the relation between stimulus and response. Or at least it seems simple.

But what about the converse – if we specify a profile of responses, then what stimulus levels would have given rise to that profile? This problem is not one that an academically-oriented psychophysicist might ordinarily face because it speaks to the issue of solving a particular problem, rather than establishing a body of knowledge. Looking at the foregoing quadratic equation, the question might be how to identify a set of ingredients for a cookie product corresponding to a given sensory profile. This type of problem-solution often arises in applied product development, and has been solved by methods published elsewhere (Moskowitz, 1999). The importance of this problem is not its contribution to a body of knowledge, but rather how psychophysical thinking both informs product development, and leads to new vistas that psychophysicists might wish to explore in the years to come.

From tongue to the mind – concept research, conjoint measurement and modeling

The business community did not limit itself to product modeling, but also reached out to psychophysical approaches by adopting conjoint measurement. Originally conjoint measurement, as proposed by Luce and Tukey in the 1960's, constituted 'fundamental measurement' based upon responses to combinations of stimuli. By the late 1960's, however, academics and practitioners in business recognized the value of conjoint measurement as well as other forms of decompositional analysis as ways to understand responses to concepts (Anderson, 1970). Concepts comprise short written vignettes about products or services. Systematically varying and testing these combinations among consumers enables the researcher to estimate the marginal or part-worth contribution of each concept element (component) to the consumer reaction.

The growth of conjoint measurement in business applications typifies applied psychophysical thinking. Whereas the aforementioned product research worked with continuous variables that could follow non-linear function forms, conjoint measurement works in the world of binaries. Researchers working with conjoint measurement want to estimate the number of rating points that are contributed by the presence of different elements.

A sense of what conjoint analysis does and what it accomplishes is shown by Figure 2 and Table 1. Conjoint analysis begins with a set of elements or components, divided into categories or silos (e.g. different product features, etc.). Experimental design systematically mixes/matches these components into small vignettes, rated by observers. The ratings at the individual level are deconstructed into the component contributions. In many cases the ratings are first transformed into a binary scale to reflect disinterest or interest in the concept, and then the analysis is run. Figure 2 shows a test concept, with the underlying architecture revealed on the left hand side (but not seen by observers in the actual test session). The actual study comprised six categories, each with six elements, or a total of 36 elements.

At the individual observer level the researcher deconstructs the combinations, using regression analysis to estimate the part-worth (marginal) contribution of the individual elements. Such analysis can be done for total panel, and for gender, as well as for segments (discussed below). The numbers in the body of the table are the additive constant (basic conditional probability of interest in the concept with no elements present), and the element utilities (incremental/decremental conditional probability of being interested in the concept if the element is added). The conjoint analysis generates a large set of utilities, one per element, for each key subgroup of interest. Key subgroups might be total panel, gender, usage pattern. Other key subgroups might emerge from an analysis of individual differences, to reveal groups of observers with similar mind-sets, but who, from all other indicators in the data, are similar in most other respects. From the analysis the researcher rapidly understands how the presence of different elements or ideas in a cookie concept or vignette drives interest. For concepts, therefore, the individual concept elements become the stimuli whose effects are studied by developing functional relations between stimulus element and observer response.

Figure 2: Example of test concept with underlying architecture revealed

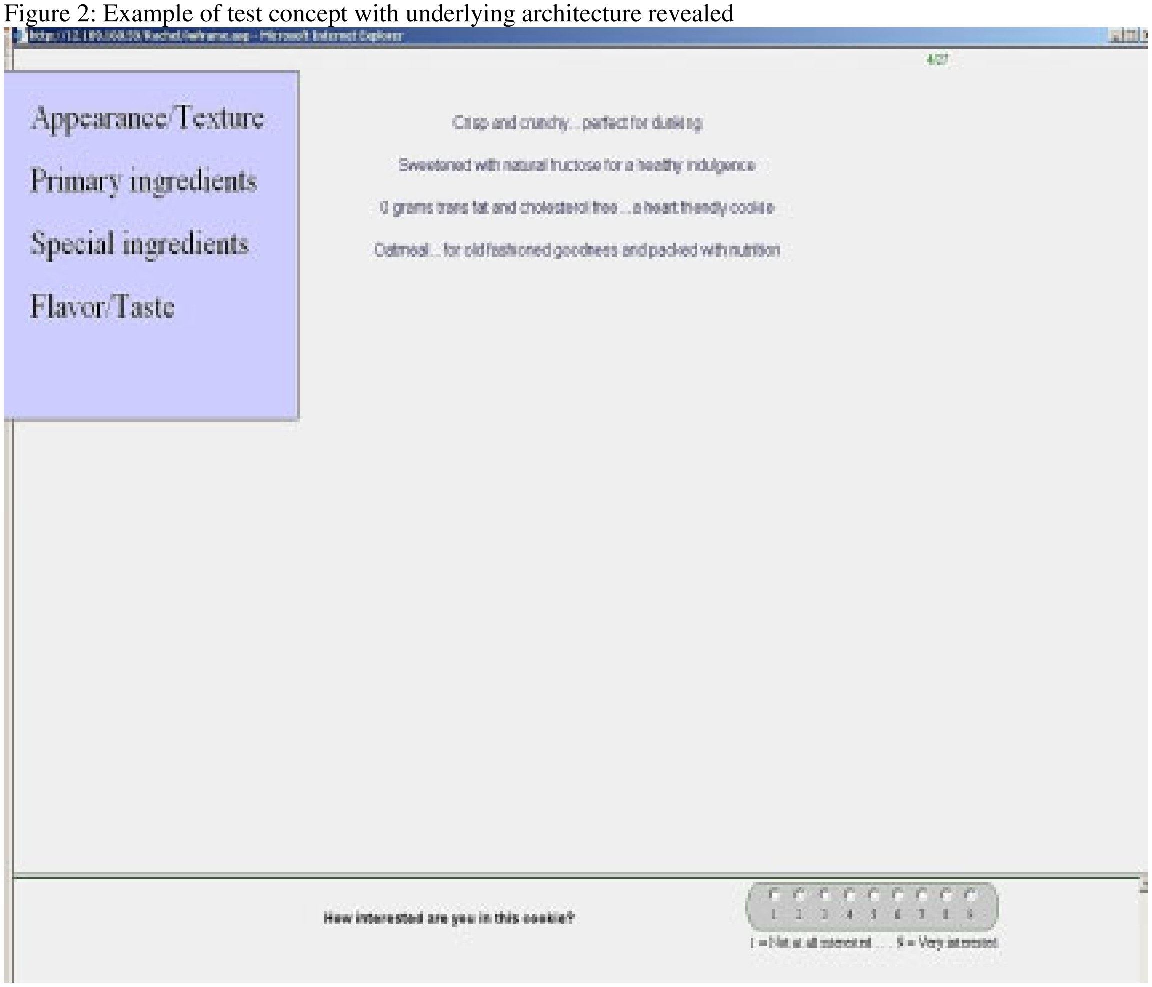


Table 1: Part-worth contributions of winning and losing individual elements. The additive constant is the conditional probability of an observer being interested in cookies without any elements. The utilities are the additive or subtractive conditional probabilities of the individual elements driving 'interested' when the elements are incorporated into the basic cookie concept.

		Gender		Concept Response Segments		
	Tot	Male	Fem	S1	S2	S3
Number of observers	439	95	344	291	59	89
Additive (basic interest in cookies)	34	34	34	37	24	33
Winning elements (total)						
Dark Belgian chocolate, Swiss milk chocolate and bittersweet chocolate…simply irresistible	**11**	**10**	**11**	**13**	**9**	4
Soft and chunky…for an extra special treat	**8**	**10**	**8**	**8**	1	**14**
Oversized chunks of dark chocolate to sink your teeth into	7	**8**	6	**12**	-3	-3
Real creamery butter for a rich, indulgent taste	7	3	**8**	8	**17**	-4
Losing elements (total)						
Vanilla flavored…a traditional favorite	-2	-6	-1	-4	-1	6
Comes in spicy flavors…cinnamon, nutmeg and allspice	-3	-9	-1	-8	3	11
Comes in cool, citrus flavors…orange, lemon and lime, perfect for a ladies afternoon tea	-9	-10	-9	-17	10	5

Conjoint analysis does not stop at main effects modeling of the type shown in Table 1, just as psychophysics does not stop at single perceived intensity curves for a stimulus that varies only in physical intensity. Researchers studying the impact of concept elements continue to demand that the conjoint methods deal with interactions, just as product modeling for foods incorporates pair-wise interactions.

During the past two years researchers have stepped up their contributions to conjoint analysis, to identify synergisms/suppressions (where the mixture performs far better or worse than expected). A great deal of the thinking behind these studies of interactions comes from the model of stimulus-response presented by psychophysics. The goal of both conjoint analysis and psychophysics is to reveal lawful relations in the data.

Interactions among elements, like interactions among variables in product research, can be estimated if the experimental design is properly set up. For products, with a limited number of independent variables, the design is no problem. For concepts, with six categories, each having six elements as a typical design, there are 15 pairs of categories, 36 pairs of elements for each category-pair, or a total of 540 combinations. The variables are integer; present or absent, rather than continuous. Estimating the combinations is difficult unless the researcher follows a strategy that differs from conventional experimental design and psychophysical thinking. If the experimental design is permuted to produce 300 or more isomorphic designs, each valid, each formally equivalent to the other, then among these 300 designs will be all pair-wise combinations of elements. Not only will there be 36 elements; the raw data will also contain all 540 pair-wise combinations, albeit spread across the different observers. This strategy, developed by Moskowitz & Gofman (2004), reveals that of the 540 pair-wise interactions possible in the cookie data presented in Table 1, only four combinations showed significant interactions. Two showed synergism and two showed suppression (see Table 2).

Table 2: Pairs of elements in the cookie study showing significant interactions.

Synergism – the combination is far stronger than expected		Effect
Oversized chunks of dark chocolate to sink your teeth into	Vanilla flavored…a traditional favorite	11
Made with canola oil which helps lower blood cholesterol levels	With added iron and isoflavones… a cookie that not only tastes good but is good for you	10
Suppression – the combination is far weaker than expected		
Soft and chewy…just like homemade	Made with canola oil which helps lower blood cholesterol levels	-9
Soft and chewy…just like homemade	Comes in a crush-proof box…no more broken cookies	-9

Some key observations and conclusions

Applied product and concept testing in industry deals with the practical problems of creating and selling products. Over the past forty years the influx of psychophysicists into industry has changed the way R&D and marketing do their jobs. Whereas once the goal was to describe the product by expert panels (hearkening back to the days of Introspection) today the strategy is to identify what works, both physically and conceptually, both in actual product and in product description.

The psychophysical heritage has not been strictly one way. Academically oriented psychophysicists have the luxury of doing experiments to build a science. The psychophysics proceeds as magisterial, often slow tempo, as the science is built and the results archived. In contrast, applications in industry have to move quickly to capture the opportunity. The psychophysics in industry is a bit rougher, cruder, but in some ways more inventive. Large scale studies with many independent variables, reverse engineering and the creation of simple but workable polynomial models holds sway in product research. Equally large scale studies with many categories and many elements per categories holds sway in concept research, where segmentation is a standard way to discover new groups of people, and where new discoveries about interactions show yet another path into the consumer's mind.

All in all, if we had to say parting words about applied psychophysics in the industrial world and its applications to products and concepts, we'd probably say *'Keep tuned .. you ain't seen nothing yet*'.

References

Anderson, N. 1970. Functional measurement and psychophysical judgment. Psychological Review, 77, 153-170.

Luce, R.D. & Tukey, J.W. (1964); Conjoint analysis: A new form of fundamental measurement; IN: Journal Of Mathematical Psychology, 1; pp. 1-36

Moskowitz, H.R. Inter-relating data sets for product development: The reverse engineering approach. Food Quality and Preference, 1999, 11, 105-119

Moskowitz, H. R., & Gofman, A. (2004). System and method for performing conjoint analysis. Provisional patent application, 60/538,787, filed January 23, 2004.

Indirect Approaches to Direct Measurement

Louis Narens
Department of Cognitive Sciences
University of California, Irvine
email: lnarens@uci.edu

abstract

Two kinds of measurement techniques have been used to measure psychological intensity: *indirect techniques,* which are based on objective observations of behavior, and *direct techniques,* which measure by subjective assignments of numbers or numerical relations. It is shown that direct measurement techniques can be reduced to indirect ones if the direct measurement *data* satisfies certain structural properties. This form of reduction provides for a rigorous theory of the scale type of the direct measurement data. This paper also uses the reduction to provide an explanation for the paradoxical empirical finding that observers in many direct measurement paradigms *behaviorally* do not distinguish differences from ratios.

Introduction and Background

Indirect measurement was introduced into psychology by Fechner (1860) who developed a theory of the shape of the psychophysical function based on the indirect measurement of thresholds. He concluded that because the threshold data empirically satisfied Weber's law, the psychophysical function must have the following form:

$$\text{subjective} = \log(\text{physical})\,.$$

Direct measurement was introduced into psychology by Plateau (1872) who developed a theory of the shape of the psychophysical function based on the equal ratios rule,

$$\frac{\text{psychological}(a)}{\text{psychological}(b)} = \frac{\text{psychological}(c)}{\text{psychological}(d)} \ \text{ iff } \ \frac{\text{physical}(a)}{\text{physical}(b)} = \frac{\text{physical}(c)}{\text{physical}(d)}\,,$$

where "psychological(x)" stands for the psychological measurement of x, and "physical(x)" for the physical measurement of x. He concluded that because his direct measurement data was consistent with the equal ratios rule, the psychophysical function must have the following form:

$$\text{subjective} = \text{constant} \cdot (\text{physical})^{\text{power}}\,.$$

From the 1930's to 1960's S. S. Stevens and others developed a number of direct measurement methodologies which they applied to a variety of stimulus domains. This tradition continues to present. From the 1960's to present, psychological researchers have developed new techniques for indirect measurement based on representing a qualitative structure by structure preserving mappings into a numerically based structure (Representational Theory of Measurement). Narens (1996, 1997, 2005) and Luce (2002, 2004) employed the Representational Theory to formulate axiomatic versions of direct measurement situations that make explicit certain critical assumptions inherent in numerically representing subjective ratio data. These axioms were tested by colleagues of Narens and Luce, and it was found that certain direct measurement techniques failed the critical axioms and others satisfied them.

This paper focuses mainly on ratio representations of direct measurement data. Such data can be collected through a variety of instructions. A commonly used instruction is, "Find a stimulus x that is p-times as intense as t." By varying t, a function f_p,

$$f_p = \{(t, x) \mid x \text{ is judged } p\text{-times as intense as } t\}\,,$$

is determined. f_p is an example of a **magnitude function.** (Other examples of eliciting instructions and their corresponding magnitude functions are provided throughout the paper.) There are two ways to represent f_p through direct measurement: A **strong form** where the stimuli is measured by a psychological function ψ such that if $x = f_p(t)$ then

$$\psi(x) = p \cdot \psi(t)\,,$$

and thus f_p is represented by ψ by multiplication by p; and a **weak form** where the stimuli is measured by a psychological function θ such that if $x = f_p(t)$ then

$$\theta(x) = k \cdot \theta(t)\,,$$

where k is some positive real number that is not necessarily p. This paper focuses on the weak form.

Continuous Ratio Production

Continuous Ratio Production is a procedure in which for each pair of stimuli a and b and for each stimulus t in X, the observer can adjust a stimuli x to satisfy the command, "Find a stimulus x so that the ratio of the subjective intensity of b to a is the same as the ratio of the subjective intensity of x to t." It is assumed that these adjustments (*i*) produce a $\preceq$-strictly increasing magnitude function R_{ab} on $\langle X, \preceq\rangle$ defined by,

$$R_{ab}(t) = x\,,$$

and (*ii*) satisfy the following three axioms:

1. **Continuum of Stimuli:** $\langle X, \preceq\rangle$ is a totally ordered continuum of stimuli.

2. **Homogeneity:** For all t and x in X there exist a and b in X such that $R_{ab}(t) = x$.

3. **Commutativity:** For all a, b, c, and d in X,

$$R_{ab} * R_{cd} = R_{cd} * R_{ab}\,.$$

The following theorem shows that Continuous Ratio Production, via the Representational Theory, yields a ratio scale. In this case, the theorem reduces direct measurement to indirect measurement by replacing the assumption that subjective ratios correspond to numerical ratios with the behaviorally observable axiom of Commutativity.

Theorem: *Suppose $\mathfrak{X} = \langle X, \preceq, R_{ab}\rangle_{a,b\in X}$ is a continuous ratio production structure. Then there exist $\mathcal{S}$ and functions S_{ab} on $\mathbb{R}^+$ such that (*i*) $\mathcal{S}$ is the ratio scale of isomorphisms of $\mathfrak{X}$ onto $\mathfrak{N} = \langle \mathbb{R}^+, \leq, S_{ab}\rangle_{a,b\in X}$, and (*ii*) for each a and b in X, $S_{a,b}$ is a function on $\mathbb{R}^+$ that is multiplication by a positive real.*

Empirical Tests of Commutativity

Various indirect axiomatizations for weak and strong forms of direct measurement have appeared in the literature, and like in the case of Continuous Ratio Production, several of these isolated the commutativity of of magnitude functions,

$$f_p * f_q = f_q * f_p \, ,$$

as the critical condition for direct measurement data to satisfy in order for the magnitude functions to be represented consistently by numerical ratios. The following studies, using various magnitude production paradigms, tested for and found the holding of commutativity of magnitude functions: Ellermeier and Faulhammer (2000), Peißner (1999), Steingrimsson and Luce (2005), Ellermeier, Narens, and Dielmann (2003), and Zimmer (2004).

Axiomatic indirect approaches has also been given for magnitude production and estimation data for the strong forms of direct measurement (e.g., Narens, 1996). Empirical tests of a critical axiom in these axiomatizations show that *strong forms* of Stevens' direct measurement methods lead to dramatic inconsistencies *if the right data are collected* (Ellermeier and Faulhammer, 2000; Peißner, 1999; Zimmer, 2004).

Torgerson's Conjecture

Torgerson (1961) discusses several experiments involving direct judgments of subjective intensity and concludes,

> These results are all consistent with the notion that the subject perceives only a single quantitative relation between stimuli. When this relation is interpreted as either a psychological distance or a psychological ratio, it can be shown that the subjective magnitudes obey the properties of the corresponding commutative group—the addition group for the distance interpretation and the multiplication group the ratio interpretation. *(pg. 205)*

I will call this conclusion *Torgerson's Conjecture,* because it is my and others' view that it does not follow from the experimental data Torgerson presented. In particular, the data were not rich enough to show that "the subjective magnitudes obey the properties of the corresponding commutative group—the addition group for the distance interpretation and the multiplication group the ratio interpretation." The following is a more detailed description of the Conjecture.

Let $R(x, y)$ is the magnitude function for the instruction, "The ratio of y to x is p," and $D(x, y)$ is the magnitude function for the instruction, "The difference between y and x is q." Let (1) and (2) be the following propositions:

(1) Subjective intensity for both ratios and differences are measured on a ratio scale, $\mathcal{S}$.

(2) $\varphi \in \mathcal{S}$ and R is represented by φ as a multiplication by a positive real.

Torgerson conjectured and found empirical support for the following:

Conjecture: If (1) and (2) above hold, then D is represented by φ as a multiplication by a positive real.

There are two problems with Torgerson's approach: (*i*) The experiments employed direct measurement techniques, and (*ii*) ratio scalability was not tested for in a rigorous manner. Other researchers have approached (*i*) by by rigorously demonstrating the related proposition, "Judgments of ratios are judgments of differences." For example, Birnbaum and Mellers (1978) showed

$$R(a, b) < R(c, d) \text{ iff } D(a, b) < D(c, d)\,.$$

However, the related proposition is not sufficient to establish (*ii*)—that the various R's and D's can be simultaneously represented by multiplications by positive reals—and thus not sufficient to establish the Conjecture as presented in Torgerson (1961).

Narens (1996, 1997, 2005) shows that the Conjecture is implied by the following proposition: **The judgments of ratios and differences are made on the same ratio scale.** His argument is based on the following theorem:

Theorem. *Suppose $\langle X, \preceq, P_j, T\rangle_{j\in J}$ is measured by a ratio scale $\mathcal{S}$ of isomorphisms onto $\langle \mathbb{R}^+ \leq, P'_j, T'\rangle_{j\in J}$, where T is a strictly $\preceq$-increasing function from X onto X. Then T' is multiplication by a positive real.*

Using this theorem, Narens shows that Torgerson's unsupported direct measurement hypothesis that R is represented by φ as a multiplication by a positive real can be deleted to yield the following indirect measurement result:

Theorem. *Let $R(x, y)$ is the magnitude function for the instruction, "The ratio of y to x is p," and $D(x, y)$ is the magnitude function for the instruction, "The difference between y and x is q."* ***Suppose:*** *R and D are measured on a common ratio scale, $\mathcal{S}$. Let φ be in $\mathcal{S}$.* ***Then:***

• *R is represented by φ as a multiplication by a positive real (one of Torgerson's hypotheses).*

• *D is represented by φ as a multiplication by a positive real (Torgerson's Conjecture).*

• *$R * D = D * R$ (the commutativity of ratios and differences).*

The commutativity ratios and differences is a consequence of Torgerson's Conjecture. The above theorem suggests that this commutativity principle provides an interesting empirical constraint. A test of it was performed by Ellermeier, Narens, and Dielmann (2003) who found that $R * D = D * R$ held for most subjects. This empirical result can be viewed as another demonstration Torgerson's Conjecture. It has the following distinguishing attributes:

• It uses a new paradigm (commutativity) to demonstrate Torgerson's Conjecture.

• Unlike Torgerson and many others, it uses indirect rather than direct methods.

• Unlike Birnbaum & Mellers, Torgerson and many others, it provides a rigorous test that it is consistent to represent subjective ratio productions in many psychophysical paradigms as numerical ratios.

In summary, the above theorem and empirical result provides the following explanation for Torgerson's Conjecture: *The collected data is representable by a ratio scale.*

Conclusions

It is important to distinguish *direct measurement data* from *direct measurement.* Direct measurement data can be used in various ways, for example, as input into a direct measurement procedure, or as in this paper, a means for testing qualitative axioms of a representational measurement theory for magnitude production.

Direct measurement data is incorporated into the Representational Theory of Measurement by showing the data satisfies specific structural properties. These properties vary with the direct measurement technique. For the kinds of paradigms thus far tested, they fail empirically for strong forms of direct measurement but hold empirically for weak forms.

The theorems presented in this paper show that the assumption of the ratio scalability of subjective intensity has powerful consequences. In particular it was shown that it implies the commutativity of ratio and difference magnitudes functions and provides an explanation of Torgerson's Conjecture.

A more complete discussion of the issues raised in this paper is presented in Narens (2005).

References

Birnbaum, M. H. and Mellers, B. A. (1978). Measurement and the mental map. *Perception & Psychophysics, 23,* 403–408.

Ellermeier, W. and Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception & Psychophysics, 62(8),* 1505–1511.

Ellermeier, W., Narens, L., and Dielmann, B. (2003). Perceptual ratios, differences, and the underlying scale. In B. Berglund and E. Borg (Eds.), *Fechner Day 2003. Proceedings of the 19th annual meeting of the International Society for Psychophysics,* 71–76. Stockholm, Sweden: International Society for Psychophysics.

Fechner, G. (1860). *Elemente der Psychophysik.* Leipzig: Breitkopf und Härtel.

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review, 109,* 520–532.

Luce, R., D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review, 111,* 446–454.

Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology, 40,* 109–129.

Narens, L. (1997). On subjective intensity and its measurement. In A. A. J. Marley (Ed.), *Choice, Decision, and Measurement: Essays in Honor of R. Duncan Luce,* 189–206. Mahwah: Lawrence Erlbaum and Associates.

Narens, L. (2005). Symmetry, Direct Measurement, and Torgerson's Conjecture. Under review, *Journal of Mathematical Psychology.*

Peißner, M. (1999). *Experimente zur direkten Skalierbarkeit von gesehenen Helligkeiten.* Unpublished master's thesis, Universität Regensburg.

Plateau, M., J. (1872). Sur la measure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bulletin de l'Académie Royale des Sciences, es Lettres et des Beaux Arts de Belgique, 33,* 376–388.

Steingrimsson, R., and Luce, R. D. (2005). Empirical evaluation of a model of global psychophysical judgments: II. Behavioral properties linking summations and productions. *Journal of Mathematical Psychology,* under revision.

Torgerson, W. S. (1961). Distances and ratios in psychological scaling. *Acta Psychologica, 19,* 201–205.

Zimmer, K. (2004). Examining the validity of numerical ratios in loudness fractionation. *Perception & Psychophysics,* in press.

DO YOU HEAR WHAT I HEAR? A STUDY OF PHANTOM TONES IN AUDITORY SIGNAL DETECTION OF FAINT TONES.

Preliminary report

Robert Z. Norman
Dartmouth College, Hanover, NH, USA
Robert.Z.Norman@dartmouth.edu

Abstract

Subjects in a 2-alternative forced-choice auditory detection experiment were asked to identify the interval in which exactly one of two known but hard-to-hear tones is presented in a background of white noise. When subjects report the wrong interval it could be because they did not hear a tone and they simply guessed the wrong interval, but it also could be that they believed they heard a tone in that wrong interval. This study looks into the source of these phantom tones. Are there components of the background noise that sound like tones, or are the tones created from within the subject? Our preliminary evidence is that neither alternative fully accounts for the data; these phantom tones arise from a significant mixture of external and internal sources.

This paper deals with auditory signal detection experiments using the two-alternative forced-choice paradigm. The signal is one of two sinusoidal tones presented in a background of white noise, and the subject is to tell which interval contains the tone. Experiments of this sort (e. g., Johnson and Norman (1), Larkin and Greenberg (2), Penner (3), Norman et al. (4)) vary in using differing probabilities of presentation of the two tones, and in payoffs to the subjects. They are concerned with how factors such as payoffs and *a priori* probabilities affect the rate of detection. Some experiments involve simultaneous presentations of the tones, and sometimes subjects are asked to identify the tones as well as the interval containing the tone. Always, one and only one of the two intervals contains a tone. In our experiments the intensities of the tones are set empirically, typically so that in the case where only one tone is presented, (either high or low) , the subject identifies the interval about 80% of the time.

When subjects are interrogated after experimental sessions about how they happened to misidentify some intervals, not infrequently they report that some of their errors in identification came as a result of hearing a tone in the interval subsequently identified as the wrong interval. A few subjects even volunteered, before we asked them anything, that they heard phantom tones. When asked what tones they heard, they were always one of the tones they were listening for.

We are trying to determine whether these phantom tones result from mistaking some component of the white-noise background for a tone, or whether the tones are internally generated. In Johnson and Norman (1), we noted that while both possibilities could be factors, a few experimenters noted that a subject might mistake part of the white noise for a signal. On the other hand, those who commented on our presentation at the ISP meeting in Coimbra (1) indicated they felt the internal process was likely predominant.

Experiment

Eight subjects took part in this experiment. All had previously engaged in detection experiments in our laboratory. All were aware of phantom tones. For each we had records of what intensities of both the high and the low tones led to 80% correct detections. We paired the subjects so as to minimize the differences in sensitivities between them.

The experiment consisted of runs of 60 trials using two pure sinusoidal tones, one high (900 hz) and one low (700 hz). These tones were familiar to the subjects, a familiarity reinforced at the beginning of each run by presenting cue tones of these two frequencies. In each trial, one of the two tones was presented in exactly one of the two intervals. Subjects were told that their objective was to detect which interval contained the tone. They were also asked to tell us, for each trial, what tone or tones they heard. We recognize that they might hear a tone in each interval and would, in that case, have chosen the interval they think most likely contained the presented tone. They reported what they heard through one of these seven replies: no tone, low, high, or, if they heard tones in both intervals, low-low, high-high, low-high, and high-low.

After each tone presentation both subjects in a pair had to answer using appropriate keys on the computer before the next trial would be presented. Both subjects were given feedback about the correctness of their interval identification. With a short practice, subjects became quite adept at responding with interval and tone identification. With just a little experience, the subjects' rate of detection of the interval was not affected by also having to identify the tone or tones they heard.

A subject who detects the interval correctly may be responding to a phantom tone. To avoid inherent complications, we analyzed only those trials in which the subject reported hearing a tone in the incorrect interval. When subjects reported hearing a tone in both intervals, we look only at the tone reported in the incorrect interval.

Experimental Expectations

In general, if the phantom tones come from the background noise we would expect the subjects to hear phantom tones on the same trials, and when they both hear a phantom tone it should be the same tone. If the phantom tones are internally generated, the trials in which the subject report phantom tones would be independently distributed, and in trials in which both subjects hear a phantom tone we would identify the same tone about half the time.

We first examined all the trials for each pair of subjects, counting for each pair the number of trials in which each subject hears a phantom tone and comparing it with the number of trials in which they both hear a phantom tone. If the phantom tones are internally generated, the occurrence of phantom tones should be independently distributed among the trials. If the phantom tones come from an external source, the white noise, we can't expect subjects to hear phantom tones on exactly the same intervals, since subjects may have used different criteria for when to report a tone. The subject with the stricter criterion would report fewer phantom tones, and so we would expect the number of common phantom tones to be the smaller of the numbers of phantom tones reported by the two listeners.

We also analyzed the trials in which both subjects reported a phantom tone. If the source is internal we would expect the subjects would report the same phantom tone only 50% of the time. If the source is external we expect them always to report the same phantom tone.

Both of the first two analyses deal with the totals of the numbers of phantom tones of all runs with each pair of subjects. A third test is to compare the distribution of trials in each run in which the two subjects report phantom tones. If the phantom tones are internally generated the trials containing their occurrences would be uncorrelated. If the tones are externally generated, we would expect these trials to be highly correlated. A problem with this test is that other factors can also affect the correlation. One such interference will occur if the subjects' detecting capabilities vary during the experiment, such as if their capabilities are greater some days than others.

Results

The results of these three tests are quite far from the expectations of either the internal or the external hypothesis alone.

The data from experimental runs with the four pairs of subjects are in Table 1. The columns labeled *phan1* and *phan2* are the number of trials containing phantom tones for the first and second subject of each pair respectively; and *both* is the number of trials on which they both reported phantom tones. The estimates for the expected number of trials both would report a phantom tone using the internal and external hypotheses respectively are labeled *int* and *ext*. We calculated *int* by multiplying *phan1* and *phan2* and dividing by the total number of trials. We calculated *ext* by summing, for each run, the smaller of the number of phantom tones for the two subjects. For pair 1, one subject reported fewer phantom tones on each run than the other. For the other pairs, our estimate may give a slight overestimate for *ext*. For the trails in which both report a phantom tone the columns *same* and *diff* show what percentage of them are the same and different phantom tones, respectively.

Table 1

Number of Phantom tones from subject pairs							
	phan1	phan2	both	int	ext	same	diff
Pair 1, 45 runs	355	931	184	122.4	355	68.5%	31.5%
Pair 2, 32 runs	299	206	97	40.5	181	81.4%	18.6%
Pair 3, 24 runs	350	386	120	93.8	299	63.3%	36.7%
Pair 4, 18 runs	93	106	25	9.1	82	56.0%	44.0%

For pairs 1, 2, and 4 the observed *both* significantly exceeds *int* if one assumes that for each subject's phantom tones the probability of *both* is a random process with probability *both*/*phan1* or *both*/*phan2*. (The data from the third pair is not statistically significant.) For the first three pairs, the percentage of *same* trials is significantly

different ($p < .05$) from the 50% predicted by the internal hypothesis, and far from the 100% of the external hypothesis. (The data from the fourth pair is not statistically significant.) Thus the internal hypothesis alone cannot account for the results.
The problems with applying third test have been noted above. Analysis using this test also seems to suggest that neither the external nor the internal hypothesis alone can account for the correlations, but the result are not statistically significant.

Conclusions

We conducted experiments using the two-alternative forced-choice paradigm using tones of low intensity in which subjects are to tell which of two intervals contains a presented tone when the tone is one of two tones familiar to the subjects. Under these conditions, subjects often misidentify the interval because they hear a tone in the other interval. In trying to determine whether these tones come from the white noise or whether they are internally generated, we find that neither source alone adequately accounts for the phantom tones reported.

We recognize that by using faint tones, we probably have increased the likelihood that a subject will misidentify the interval. If the signal intensity were to be significantly increased, it would be interesting to discover whether the findings about the mixture of external and internal sources for the phantom tones are similar to ours.

References

Johnson, E. I., and Norman, R. Z. Strategic allocation of attention in auditory signal detection, Proceedings of the 20th Annual Meeting of the International Society for Psychophysics, 2004.

Larkin, W. D., & Greenberg, G. Z. Selective attention in uncertain frequency detection. Perception & Psychophysics, 1970, 8, 179-183.

Penner, M. J. The effect of payoffs and cue tones on detection of sinusoids of uncertain frequency. Perception & Psychophysics, 1972, 11, 198-201.

Norman, R. Z, Arnold, E. H., James, A., and Meacham, A. Some Aspects of the Role of Attention in Auditory Signal Detection, II. Proceedings of the 11th Annual Meeting of the International Society for Psychophysics, 1995, 273-278.

THE LEGACY OF ABRAHAM MOLES AND ERWIN SCHRÖDINGER IN PSYCHOPHYSICS

Kenneth H. Norwich

University of Toronto
Institute of Biomaterials & Biomedical Engineering
Toronto, Ontario, Canada
k.norwich@utoronto.ca

Abstract

We examine the ideas of two scientists, A. Moles and E. Schrödinger, with respect to their impact on psychophysics. Schrödinger founded the branch of physics that we now call wave mechanics, as a means of interpreting quantum physics. Moles was interested in the applications of Shannon's new theory of information, particularly in the arts and inexact sciences. These two men had possibly never met or compared their respective ideas, and yet they share a common theme: that the world of observation is a selection from worlds forever unseen. We shall see how their thoughts can be woven into the fabric of psychophysics.

Abraham Moles, in his book entitled *Information Theory and Esthetic Perception* (Moles, 1966) made a remarkable juxtaposition. He observed that in nineteenth century Europe two mathematically simple logarithmic laws had appeared in the sciences: one in physics and one in psychophysics. Austrian physicist Ludwig Boltzmann had founded a new branch of physics that we now call statistical mechanics, and had shown that for a gas under equilibrium conditions

$$S = k \ln W \tag{1}$$

where S is the physical entropy of a gas (distinct from information theoretical entropy, which is more familiar in psychophysics) and W is the number of ways in which N gas molecules can be distributed over n states. k is now known as Boltzmann's constant. *ln* is the natural logarithm or log base e. Entropy, S, is a measure of the randomness or disorder of the molecules of the gas. This equation was of epochal importance because it permitted the mapping of microphysics (the physics of atoms and molecules) onto "macrophysics" or classical laboratory physics. It was, in fact, a turning point in the history of science. The log function is the remarkable instrument that permitted the "derivation" of the second law of thermodynamics from properties of the molecules that made up the gas. The equation is inscribed on Boltzmann's tombstone in the Zentralfriedhof zu Wien.

More familiar to psychophysicists is the seminal equation of Gustav Theodor Fechner, a German physicist, conceived in about 1860. Fechner is credited with founding the science of psychophysics. His equation, written using his own symbols, is

$$\gamma = \kappa \log (\beta/b) \tag{2}$$

where γ is "sensation magnitude", β is "stimulus value", and b is "threshold value" (Langfeld, 1912). The origin of the log function in this case can be traced to the integration of Weber's law, $\Delta\beta/\beta$ = constant, which, in turn, can be traced back to Daniel Bernoulli's principle of utility.

Moles juxtaposed these two extraordinary logarithmic equations. He felt that the thermodynamic probability, W, of gas molecules was in some way analogous to stimulus strength, β. If he were correct, then the physical entropy, S, of a stimulus would be a direct measure of its sensation magnitude.

Moles' Conjecture

In recent years I have explored Moles' conjecture, expressed more generally: *Is sensation magnitude related linearly to the physical or thermodynamic entropy of the stimulus?* To date, I have been able to answer this question positively, but for only two restricted modalities of sensation: (1) the magnitude of the taste of saltiness of aqueous solutions, and (2) the loudness of pure tones. The complete arguments are given in two papers (Norwich, 2000, 2005). For simplicity, we consider here only the gist of the idea.

Moles' conjecture for saltiness

For a dilute solution of sodium chloride, which is a strong salt, physical chemistry provides us with the following equation (see, for example, Smith 1967):

$$-\frac{S}{2R} = \ln X - \frac{S_0}{2R} \quad , \tag{3}$$

where S and S_0 are the molar entropy and the standard molar entropy of the salt solution respectively, X is the mole fraction of sodium chloride (regard it as the concentration of the salt), and R is the gas constant, 8.3144 [joule.mole^{-1}deg K^{-1}]. *Molar entropy* is the entropy per mole of solution, which is a measure of the disorder or randomness of a single mole of salt solution. The ln-function that appears in Equation (3), can be traced back to Boltzmann's Eq (1). That is, the entropy of a gas can be transposed to, or connected to the entropy of a salt in solution. Before this equation can be applied in psychophysics, a constant must be subtracted from both sides to allow for the threshold of sensation. However, we can appreciate the basic idea without using threshold. If $-S$, the negative of molar entropy, is regarded as the sensation magnitude of saltiness, then it will change linearly with the natural log of the concentration of the tasted solution, $\ln X$. That is, Fechner's law emerges directly from physical chemistry, and effectively from Boltzmann's law. We must use the negative of S to relate to sensation magnitude because as a solution becomes more dilute, entropy, S, increases, but sensation magnitude decreases. We obtain an unexpected perk from Equation (3): we can use psychophysical measurement to provide an estimate of the magnitude of the gas constant, one of the fundamental constants of physics. The threshold that is subtracted is not known precisely, and it will affect the calculation of R. The estimated value of R is accurate to within 10-15%, but does lend credence to Moles' conjecture.

Moles' conjecture for loudness of pure tones

To explore the relationship between Fechner's law for loudness and the entropy of a pure tone, we can apply the ideal gas law to the partial pressure of air across the ear drum. The equation for the molar entropy of a gas is very similar to Equation (3) above:

$$-\frac{S}{R} = \ln P - \frac{S_0}{R} \quad , \tag{4}$$

where S is the molar entropy, R is the gas constant, and S_0 is a constant. Again, we associate $-S$ with the magnitude of sensation of loudness, and we take P to be the mean pressure of the sound wave. Hence we again retrieve Fechner's law in accordance with Moles' conjecture.

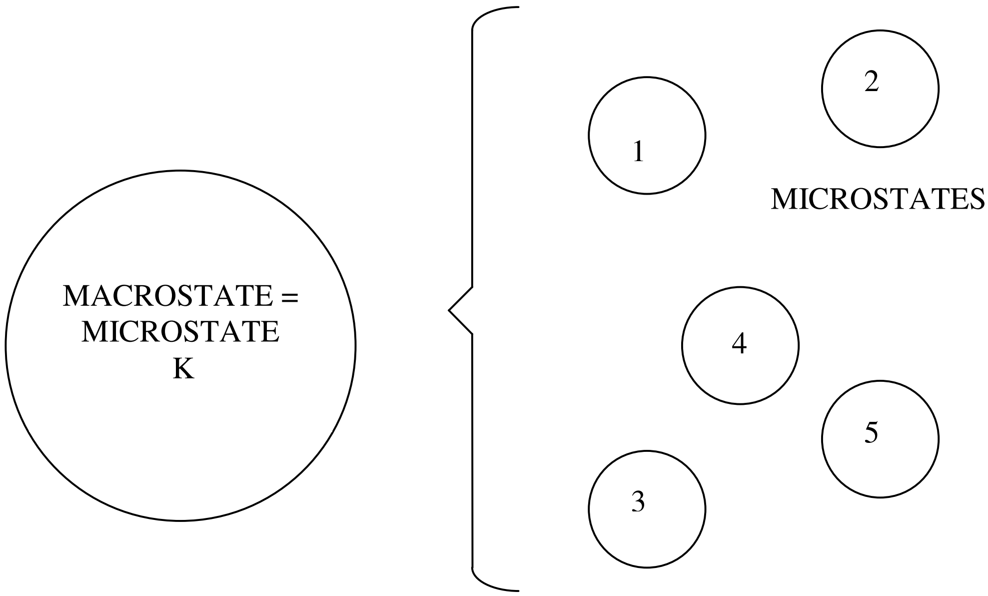


Figure 1. One macrostate is associated with many microstates. Entropy is the uncertainty about microstate *given* the macrostate.

A little mathematical manipulation will show that these logarithmic equations can also be cast into the form of power functions, which is the more usual representation of sensation magnitude today. However, I should like to continue to focus this little discussion on Fechner's law.

The concept of entropy

Beneath the surface of the above equations lurks the general principle that links psychophysics to physical entropy. Physical or thermodynamic entropy – which, again, should not be confused with information theoretical entropy – is understood as the number of ways a *macroscopic state* can be constructed out of its constituent molecules. We say that entropy is a reflection of the number of microstates in a macrostate. A macrostate or laboratory state of a system can be determined by measuring pressure, temperature, volume, etc of a system. However, there are very many ways that the (invisible) molecules that make up the system can be arranged to produce that unique combination of pressure, temperature and volume that we find in the laboratory. So at the "human" scale of measurement, which is a macrostate, we know only the net effect of molecular configurations, and we entertain a degree of uncertainty about the composition of this macrostate, which is represented by the "unknowable" microstates. Physical entropy encodes our uncertainty about these unknowable states.

Therefore, within the framework of Moles' conjecture, Fechner's law states that the magnitude of our sensation is a reflection of this underlying uncertainty. This idea is represented schematically in Figure 1.

Schrödinger's Wave Mechanics

We must now advance from the 19th to the 20th century, and change paradigms of physics, from classical to quantum physics. Actually, Schrödinger began to make contributions to psychophysics before he entered the quantum fray. His early work dealt with color theory. He studied the *threshold of distinction* for color. This seemed to be the jnd for color hue, brightness and saturation when each quality was varied independently of the other two. All colors that are at the same threshold of distinction from a given color were said to be at the same *distance* from it (Moore, 1989). I mention these matters here primarily to illustrate that physics and psychophysics were not regarded as sciences distinct from one another. However, our interest here will focus primarily on his contribution to quantum physics.

Among the achievements of Schrödinger in the 1920's was to provide a general basis for what might be called *quantum observation*, or what I prefer to call *quantum perception*. It was not that he set out to do such a thing; it emerged when he applied his new wave equation to the study of the hydrogen atom. Schrödinger regarded the hydrogen atom (which you recall consists of a single, massive proton with positive charge and a single, small electron with negative charge) in its pre-observed state, to be stationary in time. That is, he abandoned the simpler view of the electron as moving in circles or ellipses about the proton. Rather, he applied his newly hypothesized equation in the time-independent form, which looks like this:

$$-\frac{\hbar^2}{2m}\nabla^2\psi + U\psi = E\psi \tag{5}$$

ψ is known as a *wavefunction*, and different wavefunctions correspond to different energy levels of the hydrogen atom; for example, the electron could occupy certain "orbitals", each one corresponding to a unique energy level, *E*. The potential energy field is described by *U*. It transpires that there are solutions to Schrödinger's equation only for certain *E*-functions (energy eigenfunctions), which explains why the electron occupies only certain favored positions in the atom.

The physical interpretation of Schrödinger's equation and its solutions has challenged several generations of scientists. Prior to an observation's being made, the state of the quantum system as encoded by the wavefunction, ψ, is regarded as a *superposition* of all the eigenstates that are permitted solutions to Schrödinger's equation. In a sense this superposition means that the system occupies all of these states simultaneously. This strange circumstance may not disturb us unduly, because the system is still unperceived. However, as the system is observed (perceived?), the wavefunction collapses into one of its constituent eigenstates; thus in the final analysis, only one state supervenes. Before observation, the system occupies many states; after observation it occupies only one state. The situation is schematized in Figure 2.

In 1935, in order to dramatize the new concept of quantum "perception", Schrödinger put forward his well-known thought experiment now known fondly as "Schrödinger's cat". A living cat shares the space in a sealed box with a canister of poison gas, that can be opened by the decay of a radioactive nucleus. Suppose that the probability of decay is one-half. If the nucleus decays, the cat dies instantly. The unobserved nucleus exists in a superposition of states: decayed and undecayed. Hence, it would seem, in the sealed box, the cat exists also in a superposition of states: alive and dead. Only when the box is opened does the nuclear wavefunction collapse, and only then is the cat truly in one of its alive or dead states. I believe that Schrödinger was, in a way, directing our attention again to the link between physics and psychophysics. He was saying, effectively, "There is a strange intersection of quantum physics with psychophysics. Uncertainty plays a role. Let's explore it."

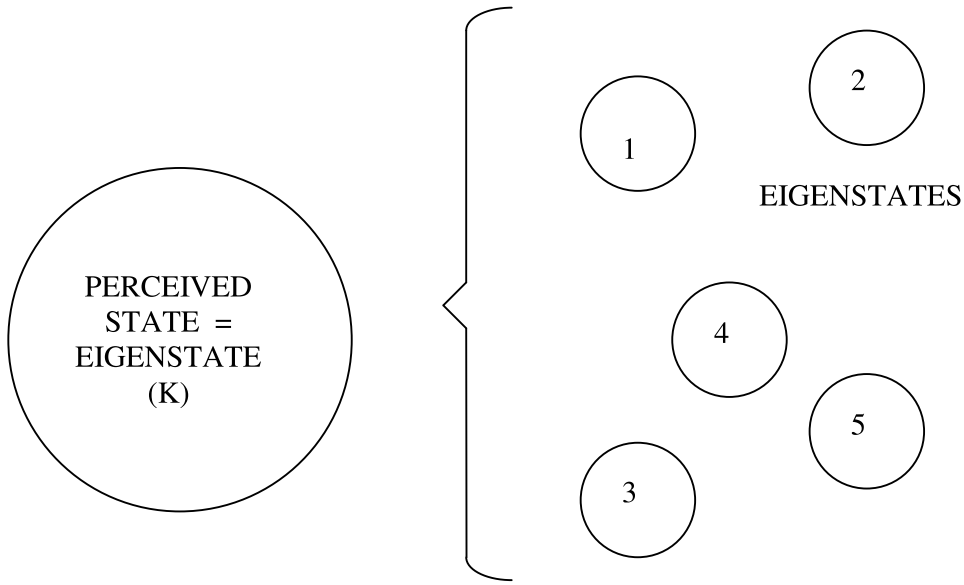


Figure 2. One observed or perceived state is associated with many eigenstates.

Evidently, the intersection of physics with psychophysics occurs at or near the point where the wavefunction collapses, and many possible states condense into one perceived state. Mathematician John von Neumann (1931) was interested in this point of intersection. He reminded us that when we observe a microscopic system, such as an atom, we use one or several instruments to render the state of the system visible to us. Then, suggested von Neumann (I am paraphrasing), the wave function of the atom is coupled with the wavefunction of the instrument to give a compound wavefunction. However, the compound wavefunction does not yet collapse. Then the atom + instrument wavefunction is coupled with the wavefunction of the human eye to produce a triply-compounded wavefunction, which also does not collapse. This wavefunction combines with that of the optic nerve etc., but where, asked von Neumann, does the chain end and the wavefunction collapse to permit perception of the microscopic event? He concluded that the compounding process terminates and collapse occurs only when human consciousness is finally achieved. Consciousness collapses the wavefunction. Much the same idea is put forth by Wigner (1983): a system collapses to an eigenstate of the observed quantity when a conscious mind engages the superposed state.

Notice that Figure 2 looks a great deal like Figure 1. In each case the perceived laboratory state, represented by the circle on the left, is a selection from the unperceived constituent states on the right. The ideas behind each of these figures evolved largely independently, the one in classical and the other in quantum physics. However, the principles involved are remarkably similar. I firmly believe that these figures encode the template for all perceptual activity.

Conclusions

All of science lies in a continuum, which we, as human beings, have seen fit to dissemble and represent as discrete units. This fragmentation serves us well in most instances, but can lead to problems at the boundaries between the artificially discrete sciences. At these boundaries, one science abuts on the other, trying to recreate the original continuum. Such may be the case for physics and psychophysics. Modern physics is progressively more concerned with the role of the observer (read "perceiver") in the process of gaining physical knowledge. Modern psychophysics, as it embraces Fechnerian principles, is borrowing, largely without being aware, from the principles of microphysics. I firmly believe that the next major advances in both these sciences will come only when each acknowledges that it must reach past the artificially set boundaries and permits itself to flow back into the original continuum.

Acknowledgements

This work has been supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

References

Langfeld, H. S. 1912. Translation of Fechner's *Elemente der Psychophysik*, in Rand, Benjamin (ed) 1912. The Classical Psychologists, pp 562-572, Boston, Houghton, Mifflin. Reprinted in Herrnstein, R. J. & Edwin G. (eds) 1965. A Source Book in the History of Psychology, pp 66-75, Cambridge, MA, Harvard University Press.

Moles, A. 1966. *Information Theory and Esthetic Perception*, pp 22-23, Urbana, University of Illinois Press.

Moore, W. 1989. *Schrödinger: Life and Thought.* Cambridge University Press.

Norwich, K. H. 2001. Determination of saltiness from the laws of thermodynamics: Estimating the gas constant from psychophysical experiment. *Chemical Senses* 26: 1015-1022.

Norwich, K.H. 2005. Physical entropy and the senses. *Acta Biotheoretica*, in press.

Schrödinger, E. 1926. Quantisierung als Eigenwertproblem (first part), *Annalen der Physik* 79, 361-376.

Smith, N. O. 1967. *Chemical Thermodynamics: A Problems Approach.* New York, Reinhold.

Von Neumann, J. 1983. *Mathematical Foundations of Quantum Mechanics*, Princeton University Press. (Originally published in German, 1931.)

Wigner, E. P. 1983. Remarks on the mind-body question, in *Quantum Theory and Measurement*, J. A. Wheeler and W. H. Zurek eds., p 177. Princeton University Press.

A FUNCTIONAL MEASUREMENT APPROACH TO THE "SELF-ASSESSMENT MANIKIN"

Armando M. Oliveira[1], Isabel B. Fonseca[2], Marta Teixeira[1], Fátima Simões[3]
[1]Institute of Cognitive Psychology – University of Coimbra, Portugal;
[2]FPCEUL-Lisbon University, Portugal; [3]University of Beira Interior, Portugal.
l.dinis@fpce.uc.pt

Abstract

The Self-Assessment Manikin (SAM) is a pictorial scale allowing for direct ratings of arousal, valence and dominance. Several of the most widely used affective stimuli database (IAPS, IADS, ANEW) have been normatively rated regarding these dimensions through extensive use of SAM's paper-pencil version. SAM is supposed to be an equal-interval scale, which is actually dependent on the way people perceive the pictorial anchors (manikins) of the scale. This issue is explicitly addressed by means of functional measurement and information integration theory (IIT) methodology. 15 subjects judged overall intensity of simultaneously presented pairs of Manikins, depicting one out of five levels of valence and one out of five levels of arousal. An equal-weight averaging integration rule was clearly established, allowing for the legitimate use of functional measurement. Functional scale values for arousal were shown to deviate somewhat from an equal-interval metric, while those for negative valence drastically departed from that assumption.

P. Lang devised the Self-Assessment Manikin (SAM) as an easy applicable, picture-oriented, scale to assess valence, arousal and dominance. According to an influential theoretical thread with roots that go back to Wundt (1896) and expressing nowadays in the several circumplex models of emotion, these are the major dimensions underlying the affective space. SAM's ratings for valence and arousal have shown almost perfect agreement with the corresponding factors obtained from the Semantic Differencial Scale developed in 1974 by Mehrabian and Russel (Bradley & Lang, 1994). However, to equivalent performance SAM adds the sizeable advantages of ease of administration (3 ratings instead of the 18 demanded by the Semantic Differential) and broader application scope (due to its non-verbal, pictorial character). This set of properties made SAM a reference scale whenever affective evaluation is construed as the rating of valence and arousal dimensions.

The paper-and-pencil version of SAM (the only one considered here) is composed of three sets of five figures (manikins), which stand for the three major affective dimensions. Figures depicting valence range from a widely smiling manikin (pleasant pole) to a low-spirited, frowning one (unpleasant pole), going through a middle neutral stance; those representing arousal range from a very relaxed, eyes-closed manikin to a highly energetic figure, displaying wide-open eyes. Finally, the dominance dimension is illustrated by figures of increasing size. Ratings of each dimension are given in a 9-point format, meaning that subjects are free to choose any of the five figures or any of the four intervals in-between them. Differences between manikins are moreover supposed to correspond to equal 2-point intervals, and the distance between any one figure and its nearest gaps to be a 1-point step. This equal-interval assumption is currently embodied in SAM's based normative ratings of widely used affective stimuli databases (IAPS, 1999; IADS, 1999; ANEW, 1999), which come expressed as means and standard deviation scores. However, it actually depends on the way subjects perceive the pictorial anchors of the scale, thus calling for an independent check.

This study is intended to provide such a check by means of functional measurement, a chapter of Integration Information Theory (IIT) (Anderson, 1981; 1982). Within IIT, measurement is not a prior to substantive inquiry, but an organic part of it - meaning, concretely, that a substantive integration rule must be first established in the concerned domain so that legitimate subjective metrics may than be derived, both for stimuli and for response (functional measurement). To achieve this, valence and arousal were taken as five-level factors (corresponding to the five manikins) in a factorial integration task asking for an overall intensity response (0-20 rating format).

Typically, the finding of an integration pattern offers simultaneous validation to the rule observed and to the linearity (i.e., equal-interval character) of the response scale (Anderson, 1981). In our case, this same response scale has actually been supported as linear in previous integration studies (concerning both emotion words and emotion faces), which helps lending credence to possible emerging patterns. Also, regarding its suitability for this specific task, there have been explicit proposals that emotion intensity arises as a joint product of valence and arousal (Reisenzein, 1994). All in all, it seems reasonably warranted that subjects are not being asked to issue an odd or unnatural judgment.

In the absence of specific predictions, any integration model (if any) might in principle occur. However, the adding rule, by its general-purpose nature, and the averaging rule, by its empirical ubiquity, could be thought off as probable candidates. Even if the averaging rule, under the equal-weight case, becomes an additive-type model, these two rules entail in fact very different properties regarding functional measurement. Of major importance is the fact that the averaging rule embodies a two-dimensional representation of scaling, allowing distinguishing weights and scale values. In case integration by average was shown the case, thus, not only the functional scale values of valence and arousal manikins could be extracted, but also their respective weights (i.e., psychological importance). Averaging would furthermore allow distances between manikins to be validly compared across factors, on a linear scale with common unit (i.e, a common interval metric for valence and arousal)

Method

Subjects. 15 graduate and post-graduate students at the University of Coimbra, all naïve regarding the scope and framework of the experiment, volunteered as subjects.

Stimuli. Digitized versions of the paper-and-pencil self-assessment manikins (obtained from the IAPS's Manual) were arranged by pairs, depicting all possible combinations of one level of valence with one level of arousal. Those pairs constituted the stimuli employed in the main factorial design. Isolated digitized manikins, standing either for a specific level of valence or arousal, were also used in the one-factor partial designs.

Design and procedure The experiment obeyed a 5 (valence) x 5 (arousal) full factorial design, to which the two one-factor subdesigns for valence and arousal were added. This was done to obtain uniqueness of weight and scale parameters estimation in case an averaging rule came to be observed (Anderson, 1982, 87-95). Each stimulus was presented twice (with reversed placement of the paired manikins) on a computer display, for a total of 70 randomized experimental trials. Subjects were run individually, after reading and commenting instructions and going through a block of 25 training trials. They were instructed to judge overall intensity conveyed by each pair of manikins on a 0-20 rating scale (with 0 and 20 treated as end-anchors, not to be used in practice), entering their response on a keyboard.

Data analysis Statistical data analysis proceeded mainly through repeated measures ANOVA, with valence and arousal as within-subjects factors.

Results

Figure 1 presents the factorial plots of valence x arousal. At observation, the immediate salient feature is an overall pattern of V-shaped parallelism, arguing for an additive-type integration rule.

Statistical analysis supports the visual inspection, by showing significant main effects on both factors ($p = .000$) and non significant interaction between them ($F(16,224) = .712$; $p = .78$), the statistical signature of parallelism.

The clear crossover exhibited by the dashed-line rising from the low-middle portion of the graph provides a crucial qualitative test between averaging (with equal weight) and adding rules, which similarly display graphic parallelism and non-significant interaction (Anderson, 1982, 215-217). It corresponds to intensity ratings given to isolated valence manikins, and it shows that while medium arousal adds something to low valence levels, it conversely takes something away from the intensity of high valence levels. This opposite effect is incompatible with an additive-summation rule, which would instead predict a parallel lower path for the dashed-line. On the statistical side, not surprisingly, valence x arousal interaction becomes significant when judgments for valence alone are included in the analysis ($F(20,280) = 8.209$; $p = .000$)

Taking evidence altogether, an equal-weight averaging model is unambiguously warranted for the integration of valence and arousal as depicted by SAM's, thereby opening way to legitimate use of functional measurement.

Functional measurement

Just as the adding rule, equal-weight averaging allows the use of marginal means of responses as functional scale values for both the row and column stimuli (the valence and arousal manikins). Given that averaging rests on a weight-value distinction, these marginal means correspond in fact to "gross stimulus values" that confound the effects of importance and valuation, and not to proper scale values – which would need extraction by procedures such as those implemented in the AVERAGE program (Zalinski & Anderson, 1986; 1991). However, this is not a problem in the equal-eight case where, due to weight constancy across the levels of each factors, "gross values" keep proportional to "proper values".

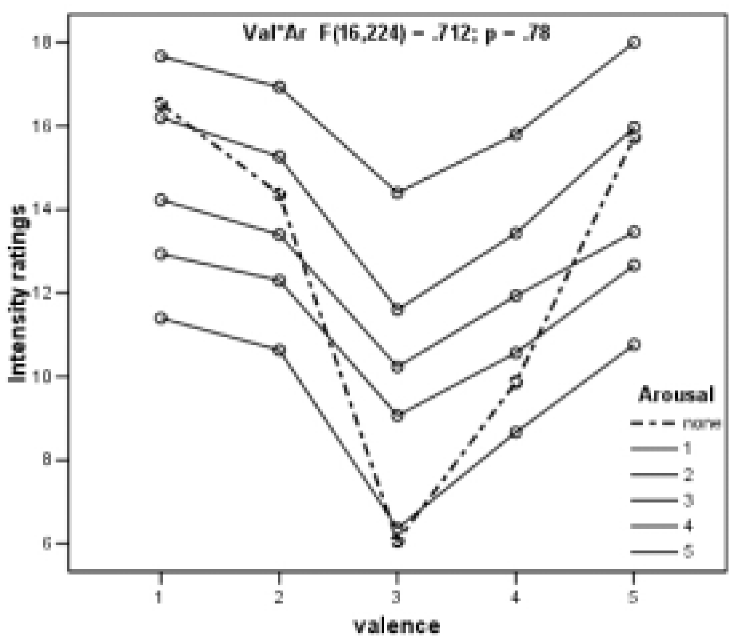
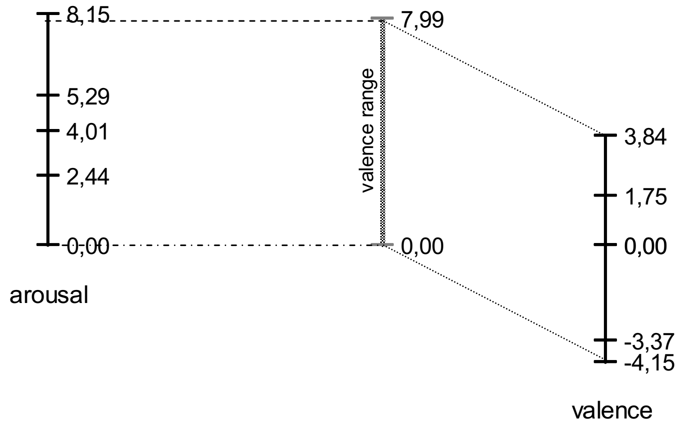
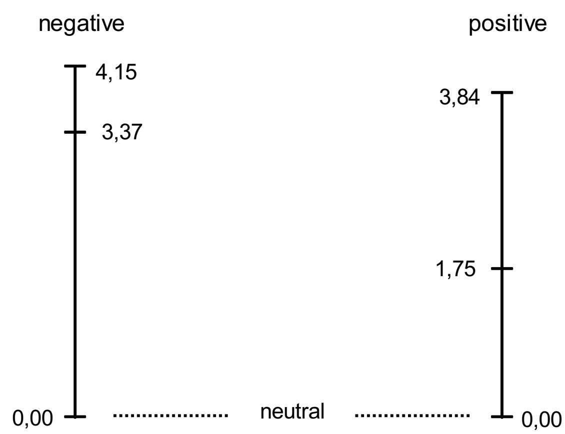


Figure 1. Integration pattern for valence and arousal: equal-weight averaging (qualitative test for averaging versus adding: dashed line)

Gross scale values for Valence and Arousal



Functional scales of Valence (gross values)



Functional scale of Arousal (gross values)

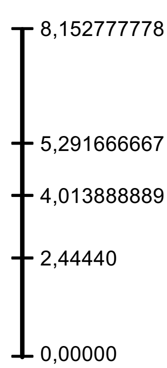


Figure 2. Functional measurement of SAM's valence and arousal

These functional values are on a linear scale, with unknown zero. Therefore, while their absolute value isn't meaningful, differences between them, differences between differences, and even ratios of differences are all sensible quantitative indices. This allows for a valid representation of the spacing between manikins, both for valence (positive and negative) and arousal, as a means to check the assumption of equal-interval spacing. Also, expressed as the ratio of distances between manikins over the total range of the factor, a weight-invariant representation of this spacing could be obtained for each factor.

Figure 2 illustrates the functional spacing between manikins as it comes out of the integration. As can be seen from the bottom and middle panel, none of the dimensions complies exactly with an equal interval metric. Arousal shows some degree of compacting of the three intermediate degrees around the mid-scale, with increased distances for the less and the more aroused manikins. As for valence, while the manikins on the positive side appear to be reasonably equidistant, those on the negative side drastically depart from that assumption, showing close vicinity of the two higher levels and a considerably larger difference between intermediate and lower levels. Conventional "zeros" were provided by the neutral manikin, for valence, and by the least excited manikin, for arousal.

At the top panel, the total ranges for each factor are matched. This couldn't be sensibly done unless weights in each factor were very similar or, indeed, the same. In fact, by using the AVERAGE program, weight estimates were shown to be almost identical for both factors: 8.22 for valence, 8.53 for arousal. Because weights were separately estimated for each subject, a paired t-test could be run over the weight distributions, providing non-significant results ($t = .646$; $p_{df=14} = .529$). Distance comparisons across valence and arousal factors can thus properly be made

Discussion

As a general outcome, the assumption of an equal-interval metric underlying SAM's valence and arousal ratings has been disavowed. The path followed to arrive at this result consisted, first, in empirically establishing an integration rule for valence and arousal as depicted in SAM - which turned out to be the equal-weight averaging model; and to induce afterwards, on basis of the model, subjective functional metrics into the stimuli variables.

Manikins standing for arousal were shown roughly equidistant for the three intermediate levels and more discrepant when the extreme levels (highest and lowest) are at stake. Manikins representing valence were almost perfectly equidistant on the positive section (starting from the neutral manikin). As for the negative side, they were drastically at odds with the equal-interval assumption, exhibiting a disproportionately large step between neutral and the first negative level, and a scarce distance between the two negative levels (even if compared with the ones registered between positive manikins)

These findings help in setting boundaries to the usefulness of SAM's paper-and-pencil ratings of affective stimuli (not just normative ones) intended for experimental manipulation. Whenever simple main effects are under focus, their use seems unproblematic. However, when interactions turn out to be of interest, SAM ratings will in general offer little support to interpretation; most of all, in no way should they be taken as the basis on which interactions are assessed. A similar limitation bears upon the common endeavour of matching "equal" positive and negative values through SAM scores (e.g., to assess the negativity bias). Or, the case being, upon attempts to go beyond the strictly monotonic level when addressing the relationship between SAM based affective ratings and physiological variables. Just as standard psychophysical functions, an extended "affective psychophysics" (Greenwald *et al.,* 1989) will need a more constraining framework than one-dimensional affective ratings to decide on the form of its functional shapes.

References

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Anderson, N. H. (1982). *Methods of information integration theory.* New York: Academic Press.

Bradley. M., & Lang. P. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry, 1*, 49-59.

Bradley. M., Lang. P., & Cuthbert, B. N. (1999). *International Affective Pictures System (IAPS): Instruction manual and affective ratings*. NIHM-Center for the study of emotion and attention, University of Florida.

Bradley. M., Lang. P. (1999). *International Affective Digitized Sounds (IADS): Instruction manual and affective ratings*. NIHM-Center for the Study of Emotion and Attention, University of Florida.

Bradley. M., Lang. P.(1999). *Affective Norms for English Words (ANEW). Instruction Manual and affective ratings.* Technical report C-1. Center for Research in Psychophysiology, University of Florida.

Greenwald, M., Cook III. E., & Lang, P. (1989). Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli.

Zalinski, J. & Anderson, N. (1989). AVERAGE: A user friendly FORTRAN-77 program for parameter estimation for the averaging model of information integration theory [computer software]. University of California, San Diego.

Zalinski, J. & Anderson, N. (1991). Parameter estimation for averaging theory. In N. H. Anderson (Ed.), *Contributions to information integration theory. Vol. I. Cognition* (pp. 353-394). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aknowledgment. This research was supported by grant POCTI/PSI/41235/2001 (FEDER/Portuguese Foundation for Science and Technology).

HEDONIC CONTRAST WITH MUSICAL STIMULI

Scott Parker, Brian Rabinovitz, Niveen Kamel
American University, Washington DC, USA
sparker@american.edu

Maya Kogan Cate Corno
Wootton HS Walt Whitman HS

Debra Zellner
Montclair State University, Montclair NJ, USA

Abstract

In two experiments subjects heard Good musical selections before Bad selections or vice versa. Good selections were better liked after Bad ones than before them. Bad selections were less well liked after Good selections than before them, but not significantly so.

Hedonic contrast was first described by Fechner in 1898 (according to Beebe-Center (1932/1965, p.222). His general principle of hedonic contrast was, "That which gives pleasure gives more pleasure the more it enters into contrast with sources of displeasure or of lesser pleasure; and a corresponding proposition holds for that which gives displeasure: (Beebe-Center p. 232, attributed to Fechner, 1898). Fechner, then, believed in both negative hedonic contrast (where stimuli preceded by better stimuli become less hedonically positive) and positive contrast (where stimuli preceded by less good stimuli become better).

Demonstrations of negative contrast are abundant (see, e.g., Zellner & Parker, 2003). Demonstrations of positive contrast are rarer. One appears in Dolese, Zellner, Vasserman, and Parker (2005) who found that unattractive paintings by Goya enhanced the ratings of more attractive paintings by Goya. Temme and Gieszen (1995) reported positive contrast with musical stimuli (traditional music was rated more pleasant after subjects listened to modern music) but didn't find negative contrast for subjects who heard the traditional music first. However, both their traditional and their modern musical selections were rated above hedonic neutrality, so they didn't have "bad" and "good" stimuli in their study.

The present two experiments attempt to demonstrate both positive and negative hedonic contrast with musical stimuli using one set of positive and one set of negative stimuli. Will presenting the bad stimuli before the good stimuli make the good stimuli better than they are when presented with no bad stimuli preceding them (i.e., will there be positive hedonic contrast)? Likewise, will bad stimuli that follow good stimuli be rated less good than when they are presented first (i.e., will there be negative hedonic contrast)? Our stimuli are musical selections, crafted to be hedonically positive or negative. We use different sets of musical stimuli in the two experiments.

EXPERIMENT 1

Method

Subjects
Forty-eight subjects were tested by three experimenters, two working at American University and the third at a high school in Rockville MD. Each experimenter tested 16 subjects.

Stimuli

The stimuli originated in ten obscure classical compositions from the 17^{th} and 18^{th} centuries. Segments from these compositions that seemed to be complete musical phrases, approximately 25 seconds in duration were downloaded from public sources. Some segments were degraded (by experimenter BR) by means of the insertion of extra rests between notes (to disrupt the rhythm) and altered pitches on some notes (to create dissonances and odd melodies). Four of the experimenters listened to the segments and rated them on a hedonic scale. We selected as stimuli five segments that had positive ratings (for "Good" stimuli) and five that had negative ratings (for "Bad" stimuli), excluding any whose ratings were so extreme as to provide a "ceiling" or "floor" for the subjects' ratings. No single composition appeared in both original form among the Good stimuli and in degraded form among the Bad stimuli.

The stimuli were recorded onto tape cassettes. Five Good stimuli with approximately 9-second spaces between them were recorded in two sequences ("forward" and "backward") on two separate tapes; five Bad stimuli were similarly recorded onto two tapes.

Procedure

Half the subjects heard Good stimuli before Bad (Group GB), and half heard Bad stimuli before Good (Group BG). "Forward" and "Backward" sequencing of Good and Bad stimuli were completely crossed with each other and with Group membership.

Subjects reported their ratings while viewing a 201-point bipolar hedonic scale with -100 marked "Hated Listening to It", 0 marked "Neither Liked Nor Disliked Listening to It" and +100 marked "Loved Listening to It".

Selections were played one at a time and the subject reported a rating which the experimenter recorded. Then the next selection was played. After all five selections on the initial tape had been played, the experimenter announced that the next tape was about to be presented. Subjects then listened to and rated the second five selections as they had on the first tape.

Results

Positive hedonic contrast will manifest itself as a difference between the ratings of Good stimuli heard first and of Good stimuli heard after Bad stimuli – higher ratings of Good stimuli by BG than by GB subjects. Negative hedonic contrast will manifest itself as a difference between the ratings of Bad stimuli heard first and of Bad stimuli heard after Good stimuli – lower ratings of Bad stimuli by GB than by BG subjects.

All subjects' average ratings of the Good stimuli were higher than their own average ratings of the Bad stimuli. Thus the subjects agreed with the experimenters about which stimuli were better and which worse.

There were no differences among the results from the three experimenters and so their three data sets were combined. There were no effects of stimulus sequence (forward vs. backward) in either stimulus set (Good or Bad) and so data from all combinations of stimulus sequences were combined.

After these simplifications, we have 48 subjects of whom 24 (Group GB) heard good selections followed by bad selections and 24 (Group BG) heard bad selections followed by good selections. If GB subjects give lower ratings to the bad selections than do BG subjects, then we have negative hedonic contrast. If BG subjects give higher ratings to the good stimuli than do GB subjects, then we have positive hedonic contrast.

Each subject's five ratings of Good selections were averaged, as were the five ratings of Bad selections. To investigate positive hedonic contrast, the two sets of 24 average ratings of Good selections were compared. The means of those averages were 21.6 for group GB and 37.3 for group BG, differing in the direction indicating positive hedonic contrast. A Mann-Whitney test comparing the two groups of averages showed them to be significantly different (U = 162.5, 413.5; z = 2.59, p<.01 two-tailed; Cliff's d = .44). Thus positive hedonic contrast occurred. Good stimuli were liked more if they were presented after Bad stimuli than if they were presented first.

In similar fashion, we compared the two sets of 24 average ratings of Bad selections. Their means were -26.8 for group BG and -39.5 for group GB, differing in the direction indicating negative hedonic contrast. However in this case the difference was not significant (Mann-Whitney U = 365, 211; z = 1.59, p = .12; Cliff's d = .27). Thus although the means differed in a way consistent with the occurrence of negative hedonic contrast, the effect was weaker than that for the Good selections if it was present at all.

Intuitively, it seems that if the Bad stimuli were more bad than the Good stimuli were good, then the Bad stimuli might be more powerful in their induction of contrast. To check if this might be contributing to our results, we compared the absolute values of the Good ratings for GB subjects against the absolute values of the Bad ratings for BG subjects. The difference between them was not significant (Mann-Whitney U = 241.5, 334.5; z = 0.96, p > .33; Cliff's d = .16). Thus differences between the amounts of observed positive and negative contrast are not attributable to differences in the initial ratings (the goodness ratings of Good stimuli versus the badness ratings of Bad stimuli).

EXPERIMENT 2

Method

Subjects

There were 48 subjects -- nineteen students at American University and twenty-nine students at Walt Whitman High School in Bethesda MD. Different experimenters ran the two groups of subjects.

Stimuli

There were six musical selections, composed and recorded by one of us (BR) onto a CD. All were simple melodic lines with few moments of harmony, approximately 25 seconds in length. Three (the Bad stimuli) had irregular rhythms, dissonances, and melodies that seemed inconsonant with western musical harmonic structure. Three (the Good stimuli) had regular rhythms and tunes that had obvious if implicit harmonies. The selections were described as being based on music from one or another third-world culture, and played on a guitar by an American musician. Both the Good and the Bad stimuli appeared in one of two sequences ("forward" or "backward").

Procedure

Half the subjects heard Good stimuli before Bad (Group GB), and half heard Bad stimuli before Good (Group BG). "Forward" and "Backward" sequencing of Good and Bad stimuli were completely crossed with each other and with Group membership.

Subjects reported their ratings while viewing a 201-point bipolar hedonic scale with -100 marked "Hated Listening to It", 0 marked "Neither Liked Nor Disliked Listening to It" and +100 marked "Loved Listening to It".

Selections were played one at a time and the subject reported a rating which the experimenter recorded. Then the next selection was played.

Results

Once again, positive hedonic contrast will manifest itself as a difference between the ratings of those Good stimuli heard first and those heard after Bad stimuli – higher ratings of Good stimuli by BG than by GB subjects. Negative hedonic contrast will manifest itself as a difference between the ratings of those Bad stimuli heard first and those heard after Good stimuli – lower ratings of Bad stimuli by GB than by BG subjects.

As in Experiment 1, all subjects' average ratings of the Good stimuli were higher than their own average ratings of the Bad stimuli. Thus the subjects agreed with the experimenters about which stimuli were better and which worse.

There were no differences among the results from the two experimenters and so their data sets were combined. There were no effects of stimulus sequence (forward vs. backward) in either stimulus set (Good or Bad) and so data from all combinations of stimulus sequences are combined here.

After these simplifications, we have 48 subjects of whom 22 (Group GB) heard Good selections followed by Bad selections and 26 (Group BG) heard Bad selections followed by Good selections. If GB subjects give lower ratings to the Bad selections than do BG subjects, then we have negative hedonic contrast. If BG subjects give higher ratings to the Good stimuli than do GB subjects, then we have positive hedonic contrast.

Each subject's three ratings of Good selections were averaged, as were the three ratings of Bad selections. To investigate positive hedonic contrast, the two sets of average ratings of Good selections were compared. The means of those averages were 22.6 for group GB and 37.7 for group BG, differing in the direction indicating positive hedonic contrast. A Mann-Whitney test comparing the two groups of averages showed them to be just about significantly different ($U = 192, 380$; $z = 1.95$, $p<.052$ two-tailed; Cliff's $d = .33$). Thus hedonic contrast occurred for the Good stimuli. As in Experiment 1, although less so, Good stimuli were liked more if they were presented after Bad stimuli than if they were presented first.

In similar fashion, we compared the two sets of average ratings of Bad selections. Their means were -22.0 for group BG and -36.9 for group GB, once again differing in the direction indicating negative hedonic contrast. As in Experiment 1, this difference was not significant (Mann-Whitney $U = 213, 359$; $z = 1.51$, $p = .13$; Cliff's $d = .26$). Thus although the means differed in a way consistent with the occurrence of negative hedonic contrast, the effect was weaker than that for the Good selections if it was present at all.

As in Experiment 1, we compared the absolute values of the Good ratings for GB subjects against the absolute values of the Bad ratings for BG subjects. Once again, the difference was not significant (Mann-Whitney $U = 259, 313$; $z = 0.56$, $p > .57$; Cliff's $d = .09$). Thus differences between the amounts of observed positive and negative contrast are not attributable to differences in the initial ratings (the goodness ratings of Good stimuli versus the badness ratings of Bad stimuli).

Discussion

Despite using two quite different types of music, the two experiments' results are quite parallel. Positive contrast emerged clearly in both studies, while negative contrast did not. The magnitude of negative contrast, while not significantly different from zero, was identical in the two experiments (Cliff's d values of .27 and .26 in Experiments 1 and 2). So for us as for Temme and Gieszen (1995), positive contrast is easier to produce and larger than negative contrast with musical stimuli. Paintings appear to be like music, in that there are to date no successful demonstrations of negative contrast. Why the pleasantness of music and paintings should differ from (e.g.) the attractiveness of birds and the goodness of beverages (stimuli which readily induce negative contrast) is not known.

References

Beebe-Center, J.G. (1965). *The psychology of pleasantness and unpleasantness.* New York: Russell & Russell. (original work published 1932).

Dolese, M.J., Zellner, D.A., Vasserman, M. and Parker, S. (2005). Categorization affects hedonic contrast in the visual arts. *Bulletin of Psychology and the Arts, 5,* 21-25.

Fechner, G.T. (1898). *Vorschule der Aesthetik II* (2nd ed.). Leipzig: Breitkopf & Hartel.

Temme, J. E. V., & Gieszen, C. A. C. (1995). Contrast effects and social desirability in art appreciation. *Empirical Studies of the Arts, 13,* 171-181.

Zellner, D. A., & Parker, S. (2003). Three influences on the eye of the beholder: Exposure, conditioning, and contrast. *Bulletin of Psychology and the Arts*, *4*, 69 - 73.

VISUAL NAVIGATION IN ADOLESCENTS WITH EARLY PERIVENTRICULAR LESIONS

Marina Pavlova [1], Alexander Sokolov[2], and Ingeborg Krägeloh-Mann [1]
[1] Developmental Cognitive and Social Neuroscience Unit,
Department of Pediatric Neurology and Child Development,
Children's Hospital, University of Tübingen, Hoppe-Seyler-Str. 1, D 72076 Tübingen, Germany
[2] Center for Neuroscience and Learning and Department of Psychiatry,
University of Ulm Medical School, Leimgrubenweg 12-14, D 89075 Ulm, Germany
e-mail: marina.pavlova@uni-tuebingen.de

Abstract

Visual navigation in familiar and unfamiliar surroundings is an essential ingredient of adaptive daily-life behaviors. Recent work helps to recognize that establishing connectivity between brain structures is of importance for successful navigation. Here we ask whether the ability to navigate is impaired in adolescents who were born premature and suffer congenital bilateral periventricular brain damage that might affect the pathways interconnecting subcortical structures with cortex. Performance on a labyrinth test was significantly worse in patients with periventricular leukomalacia as compared with premature-born controls without lesions and with term-born adolescents. The ability for visual navigation inversely relates to the severity of motor disability, leg-dominated bilateral spastic cerebral palsy. This agrees well with the view that navigation ability substantially improves with practice, and might be compromised in individuals with restrictions in active spatial exploration. Visual navigation is negatively linked to the volumetric extent of lesions over the right parieto-occipital and frontal periventricular regions. Whereas impairments of visual processing of point-light biological motion are associated with bilateral parietal periventricular lesions, navigation ability is specifically linked to the frontal periventricular lesions in the right hemisphere. We suggest that more anterior periventricular lesions impair the interrelations between the right hippocampus and cortical areas leading to disintegration of neural networks engaged in visual navigation.

Visual navigation in familiar and unfamiliar surrounding is an essential ingredient of adaptive daily-life behavior. Brain imaging data in humans along with neuropsychological evidence in lesional patients reveal a specific cortico-subcortical brain network engaged in visual navigation. Positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) repeatedly point to the right prefrontal lobe, the right hippocampus, the right medial temporal lobe, the parahippocampal gyri, the retrosplenial cortex, the medial and lateral superior parietal lobules (SPL), posterior cingulate and intraparietal cortex as parts of this distributed network (Maguire et al., 1998; Grön, Wunderlich, Spitzer, Tomczak, & Riepe, 2000). In a series of structural volumetric magnetic resonance imaging (MRI) studies, Maguire with colleagues showed that hippocampal volume positively correlates with the amount of time spent as a taxi driver in London, and concluded that visual experience in navigation plays a crucial role leading to enlargement of the posterior hippocampus (Maguire et al., 2000). In healthy adolescents and adults, similar pattern of activation involving the right frontal and anterior medial temporal lobes was found during visually guided navigation in the virtual reality (Pine et al., 2002).

Recent work suggests that connectivity between brain structures, rather than integrity of each structure per se, is of importance for successful navigating. Simultaneous lesions of several cortical areas in conjunction with the hippocampus could reproduce the impairments that thought to be observed after hippocampal lesions alone. Here, we examined visual navigation in adolescents with periventricular leukomalacia (PVL), the dominant form of brain injury in individuals who were born premature. PVL primarily affects the regions around the lateral ventricles in the peritrigonal area, and is characterized by gliosis in the white matter and tissue loss with secondary ventricular dilatation, thereby impinging on the pathways interconnecting subcortical structures with cortical regions.

Patients suffering bilateral damage to the periventricular brain regions are often reported by their care-providers, family members and clinicians to have specific difficulties in way-finding and visual navigation, and easily getting lost in unfamiliar surroundings. However, these behavioral deficits and underlying them functional brain neuropathology have not been systematically investigated. Former preterms with PVL often exhibit signs of motor disorders in a form of leg-dominated bilateral spastic cerebral palsy (BS-CP) ranging from slight signs of motor disorders through severe impairments in walking ability to complete disability for autonomous locomotion.

In the present work, within the context of a basic study aimed at assessing perceptual deficiencies in adolescents who were born premature (Pavlova, Staudt, Sokolov, Birbaumer, & Krägeloh-Mann, 2003; Pavlova et al., 2005), we ask (i) how, if at all, visual navigation is impaired in patients with PVL; (ii) whether these deficits are specifically related to the topography and extent of periventricular lesions; and (iii) whether the visual navigation ability is related to motor disability. Because it is established that the ability to navigate substantially improves with practice (Maguire et al., 2000; 2003; Winocur, Moscovitch, Fogel, Rosenbaum, & Sekeres, 2005), we expect that it might be compromised in individuals with restrictions in active locomotion leading to difficulties in spatial exploration of the environment.

Method

Participants

Patients were eleven adolescents (aged 13-16 years, mean 15, ±SD 1.19) born premature between 27 and 33 weeks of gestation with magnetic resonance imaging (MRI) evidence for PVL. Eight children who were born premature and eight term-born participants had MRI scans without any identifiable signs of brain damage or other abnormalities, and served as controls. Participants who were born premature were recruited on a voluntary basis from a data pool of the Department of Pediatric Neurology and Child Development, Children's Hospital, University of Tübingen. Term-born controls were recruited as volunteers from the local community. All participants had normal or corrected vision. Verbal IQ greater than 85 (HAWIK-III based on the WISC III, adapted to the German population) was an inclusion criterion for all participants. With respect to locomotion ability, the patients with PVL ranged from normal function through impairment in walking pattern to complete walking disability. More specifically, in eight of 11 patients with PVL, a leg-dominated bilateral spastic cerebral palsy (BS-CP) was diagnosed. All patients attended a mainstream school with the exception of one male patient who attended a special school for motor disabled children.

Labyrinth Task

The present study is focused on performance on a Labyrinth (LA) test administered to participants during neuropsychological examination (HAWIK-III). This test is aimed at assessing the ability for visuospatial navigation, topographical skills and way finding. The LA test consists of 10 tasks increasing in their complexity (Figure 1). In each task, a participant has to find the way out from the center of a 2-D maze. When assessing performance on the task, both accuracy and time needed for way finding are taken into account. Row data are then transformed into the normalized scores ranging from 1 (floor performance) through 10 (normal performance for this age) to 19 (ceiling performance). Although the patients were almost free from motor disorders of upper limbs, we excluded consideration of motor component while passing the maze (e.g., touching “the labyrinth’s walls” by a pencil) when assessing performance on the task.

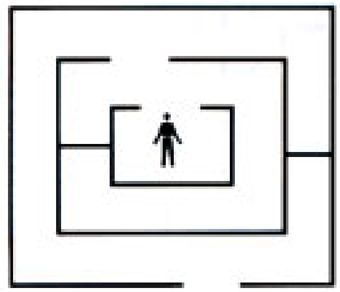


Figure 1. Examples of the 2-D labyrinth tasks administered to participants: one of the simplest labyrinths.

Results

Visual navigation ability

The performance level of patients with PVL was significantly lower than in both control groups (mean score 4.81 (SD 4.205), 9.375 (SD 3.276), 10.125 (SD 3.099), for patients, premature-born and term-born controls, respectively; t-test, two-tailed, $p < 0.04$, comparison with premature-born controls, $p < 0.02$, comparison with term-born controls). The performance scores of 7 out of 11 patients were below the normal range. There were no significant differences in performance between both control groups. This indicates that poor performance on visual navigation task in patients with PVL is not simply due to brain prematurity alone.

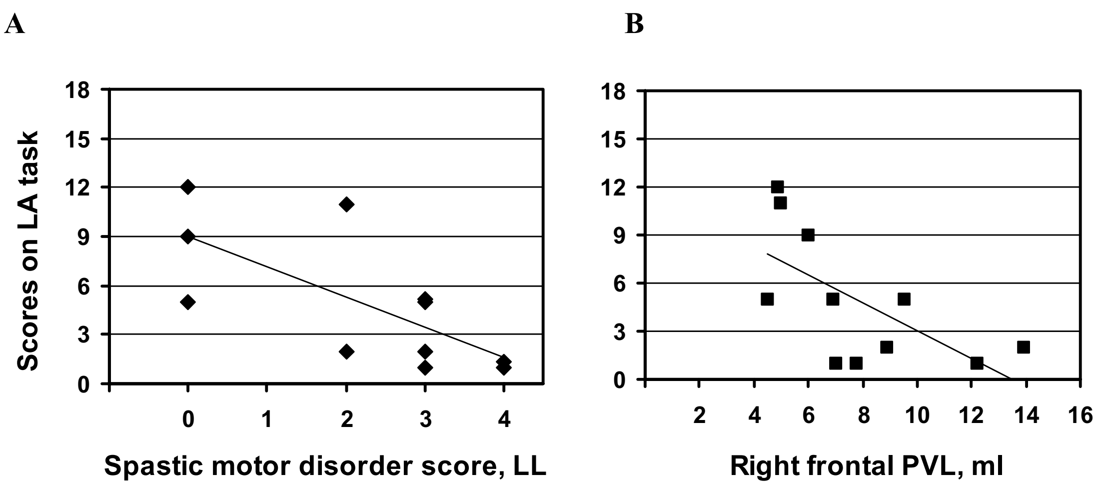


Figure 2. The scores on the LA test in patients with PVL plotted against **(A)** the spastic motor disorder scores for lower limbs, LL (r = -0.697, p< 0.05); and **(B)** the volumetric extent of PVL in the right frontal regions (r = -0.641, p < 0.05).

Visual navigation and PVL extent

In patients, performance on the LA task correlates negatively with the volumetric extent of PVL over the right parieto-occipital complex (Pearson product-moment correlation, r = -0.617, p < 0.05):

the performance level drops with an increase of the lesion extent. No relationship occurs between performance on the LA task and the lesion extent in the temporal region ($r = -0.111$, $r=0.305$, n. s., for the right and left hemisphere, respectively). The strongest correlation was found between performance on the LA task and the extent of PVL in the right frontal regions ($r=-0.64$, $p< 0.05$; Figure 2B). In earlier work, we reported a strong negative link between the extent of bilateral parieto-occipital PVL and performance on the task requiring visual processing of point-light biological motion displays representing human locomotion (Pavlova et al., 2003; 2005). We also established the inverse relationship between the extent of parieto-occipital PVL and the scores on IQ factors Perceptual Organization (PO) and Processing Speed (PS) (Pavlova et al., 2005). The specificity of the link between the severity of the right frontal and parieto-occipital PVL and the navigation ability is also revealed by the fact that there is a lack of relationship between the white matter volume, which is often considered a distinct feature of PVL severity, and performance on the LA task ($r=0.193$, n.s.).

Visual navigation ability and IQ factors

No linkage was found between performance of the patients on the LA task and the scores on either general ($r= 0.431$, n.s.) or verbal IQ ($r= 0.16$, n.s.). Most important for the purpose of the present work, no substantial link occurs between the visual navigation ability and performance IQ ($r= 0.415$, n.s.). Moreover, neither performance on Perceptual Organization tasks (IQ factor PO based on four visual tasks, namely, picture completion, event arrangement, block design, and object assembly) nor on visual attention tasks (IQ factor Processing Speed, PS, based on two tasks, visual search and coding) was substantially related to performance on the LA task. This also underscores the specificity of the impairments of the visual navigation ability in patients with PVL.

Linkage between visual navigating to locomotion ability

We further tested the hypothesis that if early restrictions in active locomotion were related to deficits in visual navigation ability, then the performance level on the LA task would decrease with an increase in the severity of motor disorders. We were mainly interested in this relationship because of two decisive reasons. First, all patients in the present study had leg-dominated motor disorders. Second, there is evidence that restrictions in active exploration of environment are essential for development of the navigation ability (Wang & Spelke, 2000). As expected, in the patients with PVL, a strong relationship between performance on the LA task and the severity of functional motor disorders of lower extremities was found ($r = -0.697$, $p< 0.05$; Figure 2A).

The data were submitted to a stepwise multiple regression analysis, with the dependent variable *navigation ability* and three independent variables, namely, *motor disability*, *extent of right parieto-occipital PVL,* and *extent of right frontal PVL*. The motor disability was entered first and explained a significant percentage of the variance (49%) in the navigation ability ($F(1;9) =8.491$, $p < 0.017$). The extent of lesions over the right frontal regions was entered second and explained an additional 25 % of variance ($F(1;7)=7.409$, $p < 0.026$). The third variable (extent of lesions over the right parieto-occipital region) did not explain a further significant proportion of the criterion variance, and could not be considered as an independent predictor of the navigation ability. We suggest, therefore, that the right frontal and parieto-occipital PVL affects the same neural network subserving visual navigation ability.

Discussion

The main outcome of the present work indicates that the visual navigation ability is profoundly compromised in patients with periventricular leukomalacia. Moreover, the severity of this impairment is specifically related to the topography and extent of periventricular damage: the performance on navigation task is inversely related to the volumetric extent of damage to the frontal and parieto-occipital regions in the right hemisphere.

Damage to periventricular regions that contain many interconnecting fibres projecting to the cortex might interrupt functioning of the distributed cortical network subserving visual navigation and way finding. Periventricular lesions might break the reciprocal thalamocortical interrelations impinging on posterior thalamocortical fibers. Recent diffusion tensor imaging (DTI) findings suggest that PVL affects the posterior thalamic radiation (Hoon et al., 2002), which connects the pulvinar and lateral geniculate nucleus (LGN) to parietal cortex. Interruptions of this connection may affect performance on a number of visual perceptual and attentional tasks. In earlier work (Pavlova et al., 2003; 2005), we showed that deficiencies in visual processing of point-light displays representing human locomotion are associated with the severity of parietal periventricular lesions in both hemispheres. We also established the inverse relationship between the extent of parieto-occipital PVL and the scores on factors Perceptual Organization and Processing Speed constituting performance IQ (Pavlova et al., 2005). Here, we found that navigation ability is linked to the parietal and, in particular, to the frontal periventricular damage to the right hemisphere. Bearing in mind that the brain network subserving visual navigation involves the right prefrontal lobe and hippocampus (Maguire et al., 1998; Grön et al., 2000), we suggest that more anterior lesions lead to disintegration in the functioning of this network breaking the interrelations between the right hippocampus and frontal cortical regions. Moreover, as indicated by the outcome of the stepwise multiple regression analysis, the extent of right frontal lesions may be considered as a predictor of the impairments in navigation ability.

The findings provide evidence in favour of, and further elaborate, the notion that it is the brain connectivity rather than integrity of a single brain structure that constitutes an essential requirement for successful visual navigation (Aguire & D'Esposito, 1999; Maguire, 2001; Holscher, 2003). This view also corresponds to the findings showing that oscillatory theta brain activity that is considered a putative indicator for interconnections between neural networks, exhibits specific enhancements during visual navigation tasks (Kahana, Sekuler, Caplan, Kirschen, & Madsen, 1999). Notably, theta oscillations occur more frequently in more complex mazes, visual navigation in which might involve more complex neural networks.

The other essential finding of the present study is that the visual navigation ability is inversely related to the severity of motor disorders, leg-dominated BS-CP. Previously, we reported that visual processing of point-light displays representing human locomotion does not depend on the ability to produce biological movement challenging the idea that motor experience is an obligatory prerequisite for the perception of human locomotion (Pavlova et al., 2003). Motor experience does not appear to be necessary for visual analysis of human movement presumably because a hard-wired schema for biological motion processing is inherent for the brain. By contrast, the present data highlight the role of active spatial exploration for a proper development of visual navigation. This outcome agrees well with the findings demonstrating that the ability to navigate substantially improves with practice (Maguire et al., 2000; 2003; Winocur et al., 2005), and might be compromised in individuals with restrictions in active spatial exploration. Like insects and rodents, humans continuously update their spatial representation of the environment as they move (Wang & Spelke, 2000). In patients with meningomyelocele (MMC) and hydrocephalus, a tight relationship is reported to occur between locomotion ability and scores on performance IQ that is based on a number of visual perceptual tasks (Rendeli et al., 2002). The lack of a link between the severity of leg-dominated BS-CP and performance IQ in patients with PVL in the present study points to the specificity of connection between active spatial exploration and the visual navigation ability in this sample of patients. Moreover, as revealed by the stepwise multiple regression analysis, the severity of locomotion disability may be considered a predictor of deficiencies in visual navigation ability.

Acknowledgements

We thank the participants and their family members for their kind cooperation. The work was supported by the Deutsche Forschungsgemeinschaft (DFG, KR 1316/5-P) and the University of

Tübingen Medical School (*f*ortüne-Programs 716 and 979, M.P.). We appreciate Franziska Hösl for technical assistance in MRI data acquisition, Martin Staudt for helping with MRI data analysis, Karen Lidzba and Arseny Sokolov for assistance in neuropsychological testing, and Christel Ildei, a visiting PhD student in MP's laboratory, for assistance in data analysis.

References

Aguirre, G. K., & D'Esposito, M. (1999). Topographical disorientation: a synthesis and taxonomy. *Brain,* 122, 1613-1623.

Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nature Neuroscience*, 3, 404-408.

Holscher, C. (2003). Time, space and hippocampal functions. *Reviews in Neurosciences,* 14, 253-284.

Hoon, A. H., Lawrie, W. T., Melhem, E. R., Reinhardt, E. M., van Zijl, P. C. M., Solaiyappan, M., Jiang, H., Johnston, M. V., & Mori, S. (2002). Diffusion tensor imaging of periventricular leukomalacia shows affected sensory cortex white matter pathways. *Neurology,* 59, 752-756.

Kahana, M. J., Sekuler, R., Caplan, J. B., Kirschen, M., & Madsen, J. R. (1999). Human theta oscillations exhibit task dependence during virtual maze navigation. *Nature*, 399, 781-784.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: A review of lesion and neuroimaging findings. *Scandinavian Journal of Psychology*, 42, 225-238.

Maguire, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S., Frith, C. D., & O'Keefe, J. (1998). Knowing where and getting there: A human navigation network. *Science,* 280, 921-924.

Maguire, E. A., Gadian, D. G., Johnsrude, I. S., Good, C. D., Ashburner, J., Frackowiak, R. S. J., Frith, C. D. (2000). Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences of the USA*, 97, 4398-4403.

Maguire, E. A., Spiers, H. J., Good, C. D., Hartley, T., Frackowiak, R. S., & Burgess, N. (2003). Navigation expertise and the human hippocampus: a structural brain imaging analysis. *Hippocampus*, 13, 250-259.

Pavlova, M., Staudt, M., Sokolov, A., Birbaumer, N., & Krägeloh-Mann, I. (2003). Perception and production of biological movement in patients with early periventricular brain lesions. *Brain*, 126, 692-701.

Pavlova, M., Sokolov, A., Staudt, M., Marconato, F., Birbaumer, N., & Krägeloh-Mann, I. (2005). Recruitment of periventricular parietal regions in processing cluttered point-light biological motion. *Cerebral Cortex*, 15, 594-601.

Pine, D. S., Grun, J., Maguire, E. A., Burgess, N., Zarahn, E., Koda, V., Fyer, A., Szeszko, P. R., & Bilder, R. M. (2002). Neurodevelopmental aspects of spatial navigation: a virtual reality fMRI study. *Neuroimage*, 15, 396-406.

Rendeli, C. V., Salvagio, E., Sciascia Cannizzaro, G., Bianchi E, Caldarelli M, & Guzzetta F. (2002). Does locomotion improve the cognitive profile of children with meningomyelocele? *Child's Nervous System*, 18, 231-234.

Spiers, H. J., Burgess, N., Maguire, E. A., Baxendale, S. A., Hartley, T., Thompson, P. J., & O'Keefe, J. (2001). Unilateral temporal lobectomy patients show lateralized topographical and episodic memory deficits in a virtual town. *Brain*, 124, 2476-2489.

Wang, R. F., & Spelke, E. S. (2000). Updating egocentric representations in human navigation. *Cognition*, 15, 215-250.

Winocur, G., Moscovitch, M. J., Fogel, S., Rosenbaum, R. S., & Sekeres, M. (2005). Preserved spatial memory after hippocampal lesions: effects of extensive experience in a complex environment. *Nature Neuroscience*, 8, 273-275.

INFLUENCE OF EMOTION CATEGORY AND INTENSITY ON AFFECTIVE BRAIN PROCESSING: CNS AND ANS INDICES

Telmo Pereira*; Isabel B. Fonseca**; Armando M. Mónica***
* Superior Health School-Coimbra, Portugal ** Psychology Faculty-Lisbon University, Portugal
*** Institute of Cognitive Psychology-University of Coimbra, Portugal
Huc_hamlet@yahoo.com

Abstract

This study endeavors to investigate hemispheric asymmetries in the perception of facially expressed emotion, as a function of emotion category and of conveyed intensity of expressions. It conjoins a 4 emotion x 2 intensity x 2 hemifield full factorial design with the collecting of behavioral and psychophysiological data. A 2AFC task was used, requiring a decision over the location (left/right) of emotion-conveying faces on presented neutral-emotional pairs. Outcomes support the approach-withdrawal view of hemispheric activation at the high intensity levels (left hemisphere privilege for approach-related affect, such as joy and anger; reversed right privilege for withdrawal-related emotions, such as fear and sadness), whereas at low intensity levels overall right dominance was instead observed. Emotional intensity thus stands out as an important determinant of hemispheric lateralization patterns, entailing the need to give it explicit consideration in any attempts to understand affective brain processing.

Human brain is, both anatomically and functionally, asymmetric. Evidence has accumulated that this asymmetrical functioning extends to emotions, particularly over prefrontal cortical territories. This has been for a long time now a core topic of scientific literature, with vast bodies of evidence converging towards one or another of the two prevailing views of emotional asymmetry - the *right hemisphere theory* and the *valence theory*.

The *right hemisphere theory* argues for an overall privilege of the right hemisphere for every form of expression/perception of emotions or, more generally stated, affect. It is buttressed by a large number of clinical as well as experimental results. Lesions in the right temporal and parietal cortices have been shown to impair emotional experience and arousal (e.g., Heller, 1993). Observations conducted with brain damaged patients appear consistent as well with right hemisphere dominance regarding emotion perception (Borod *et al.*, 1998). On the side of emotion expression, it has also been documented that emotions are more intensely expressed on the left side of the face, which suggests a stronger involvement of the contralateral right hemisphere in their facial display (e.g., Sackheim *et al.*, 1978).

Valence theory advocates, on its turn, a pattern of hemispheric lateralization dependent on emotional valence. Specifically, positive emotions are believed to associate with larger activation of the left hemisphere, while negative ones conversely correlate with augmented activation in the right hemisphere. Extended scientific support to this alternative view is also available. As an illustration, shorter reaction times for positive emotions (e.g., happiness) presented to the right-visual-field and conversely longer latencies for negative emotions (e.g., sadness) have been demonstrated in studies resorting to the divided-visual-field technique (e.g., Reuter-Lorenz & Davidson, 1981) - a result that favors a right hemisphere privilege for negative emotions and a left privilege for positive emotions. More recently, this theory underwent an important evidence-based revision, consisting in promoting an approach-withdrawal account, rather than a valence one, of hemispheric asymmetries (Davidson RJ, 2004). According to this up-dated version, approach-related emotions (e.g., positive emotions plus anger) are preferentially represented at specific left-sided prefrontal sites, while withdrawal-related negative affect is preferentially represented at right-sided prefrontal territories (Davidson RJ, 2004; Harmon-Jones E et al, 1998; Tomarken AJ et al, 1990).

Besides CNS responses, autonomic responses have also been called upon in connection with the study of emotions. However, the issue of laterality has not been at the front of this line of inquiry, more concerned with the long-lasting (since W. James) subject-matter of autonomical emotion specificity.

The aim of this study is twofold: 1) to investigate hemispheric asymmetries in the perception of facially expressed emotions as a function not only of emotion category but also of conveyed intensity; 2) to cross-bridge evidence regarding lateralization patterns arising out of behavioral, CNS and ANS responses in the framework of a common experimental design. A discrete conception of emotions was operatively adopted (e.g., Ekman's taxonomy of basic emotions) and, in order to selectively address each of the brain hemispheres, the divided-visual-field technique was employed in all stimuli presentations.

Method

Subjects 15 undergraduate students (9F, 6M; mean age: 22 $\pm$ 2,9 years) volunteered to participate in the experiment. All of them were right-handed and possessed normal or corrected to normal vision.

Stimuli and design: Pictures of neutral and emotion conveying faces selected from databases JACFEE and JacNEUF (Matsumoto & Ekman, 1988). Faces of 8 Caucasian individuals (4M, 4F) expressing joy, anger, fear and sadness at high intensity levels (assessed by normative ratings provided with the database), as well as their corresponding neutral expressions were used in the experimental trials (Japanese faces were used in the training period). Intermediate levels of intensity were obtained by morphing at equal 33% steps between neutral and high intensity expressions (with Morpheus Software). The first morph, at one third of the range, was selected to represent low intensity of a given emotion. Stimuli pairs composed of a neutral face and either a high or a low intensity emotion expression were then built for each individual and for each emotion Each pair was randomly presented four times, the emotional face being projected twice on each hemifield, for a total of 64 experimental trials. This way, the experiment embodied a 4 emotion (joy, anger, sadness, fear) x 2 intensity (high, low) x 2 visual hemifields (RVF, LVF) full factorial design.

Procedure. Participants sat in a recliner in front of a VGA monitor, in a dimly lit room. Line sight was leveled to the center of the display and observer distance was set to 50 cm, allowing to achieve the visual angles needed to implement the divided-half-field technique (e.g., around 10^0 subentending the separation between both faces, equidistant from the fixation point). Each pair of faces was displayed for 150 ms. Observers' task was to determine the side on which the face seemed to show greater emotional intensity. All aspects of presentation were managed via SuperLab 2.01, which also triggered the recording system.

EEG and ANS data collection assemblage: Six EEG leads (locations F3, F4, T3, T4, P3, P4 on the 10-20 IS) were used, all referenced to Cz. Ag-AgCl electrodes were used. Data were collected at a sample rate of 150 Hz through EEG 100B Biopac amplifiers with a band-pass filter of 0.1-35 Hz.
Autonomic responses recorded were ECG, pneumogram and electrodermal activity. ECG monitoring was done by means of three Biopac LEAD 100S electrodes, placed on the left arm, right arm, and leg (ground), according to a Lead I setting. Pneumographic recording used a Biopac TSD 101C respiration transducer, placed around the thoracic midline. Electrodermal activity (GSR) was registered with Biopac TSD 103A electrodes, set around the tip of the index and middle fingers. Autonomic data were collected at different sample rates: 75Hz (GSR and ECG) and 150 Hz (Pneumogram) .

Data Analysis: After visual inspection for artifacts, waves were edited according to each of the 16 experimental conditions defined by the 4 x 2 x 2 factorial design. Each time epoch included a 2 s baseline period and extended for 10 s after stimuli onset.

ECG and pneumographic recordings were used to derive sinus heart arrhythmia according to the peak/valley estimation method (Grossman, P., 1992). As for GSR, maximum peak amplitude (adjusted to baseline) was considered. EEG spectral analysis was performed via fast fourier transform over the first second following stimuli onset, and α-power was estimated (mean value on the α-band 8.0 – 13.0 Hz). Asymmetry was calculated according to the formula: Log (right α-power) - Log (left α-power) (Davidson, R.J. et al, 1990).
Statistical data analysis consisted for the most in repeated-measures ANOVAs (within subject factors: emotion, intensity, visual field) run over the different response measures.

Results

Fig. 1 plots results for behavioral measures, accuracy (%) and RT (ms). No significant differences in accuracy were found between hemifields. As for reaction times, however, a difference concerning visual fields emerged at the high intensity level ($p < .05$), displaying a privilege of the left visual field (LVF) for fear and sadness and a symmetrical privilege of the right visual field (RVF) for joy and anger – crossed pattern, revealing a disordinal interaction, at the bottom of the left panel.

This finding concurs with a lateralization effect along an approach-withdrawal axis. Among the pool of emotions considered, sadness presents the most distinctive profile, displaying both the lowest accuracy values and the higher reaction times.

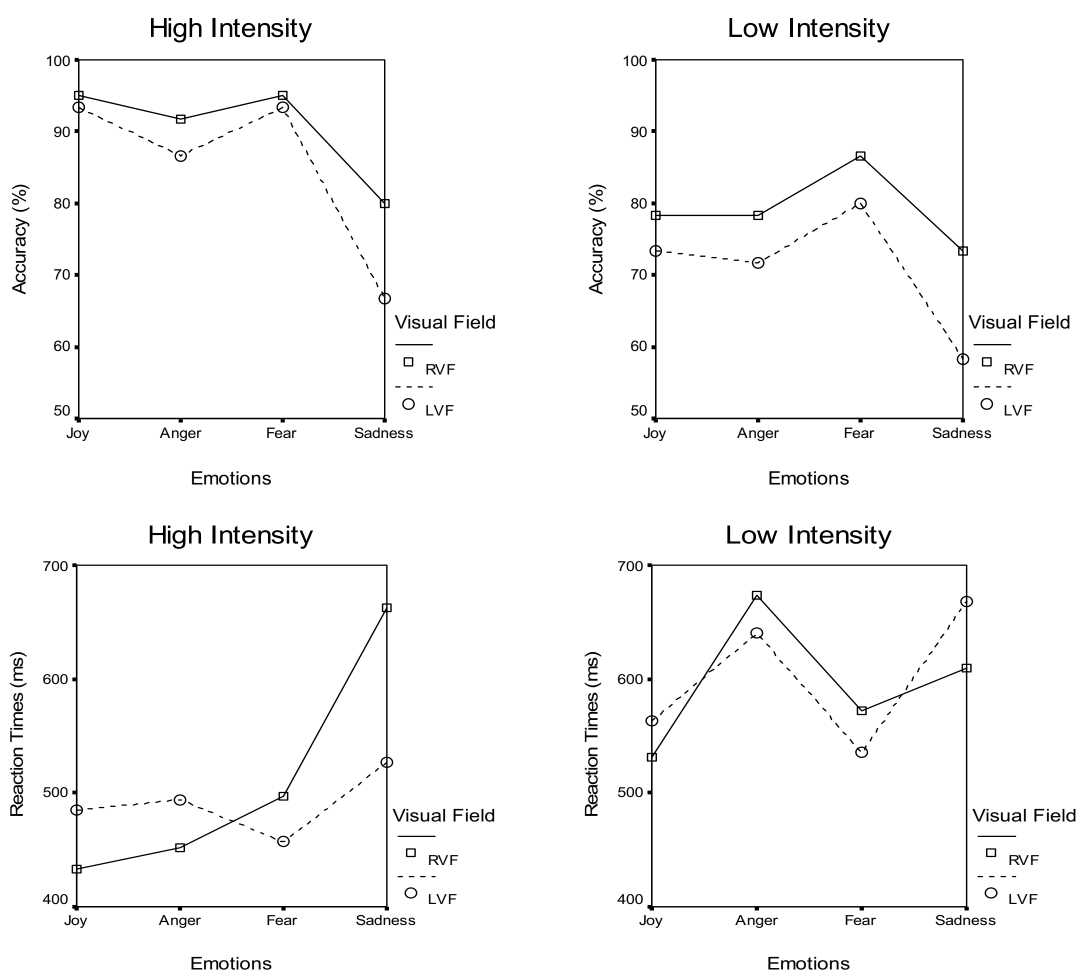


Figure 1. Factorial plots for accuracy (%) and RT (ms) (Emotion x Intensity x Visual Field)

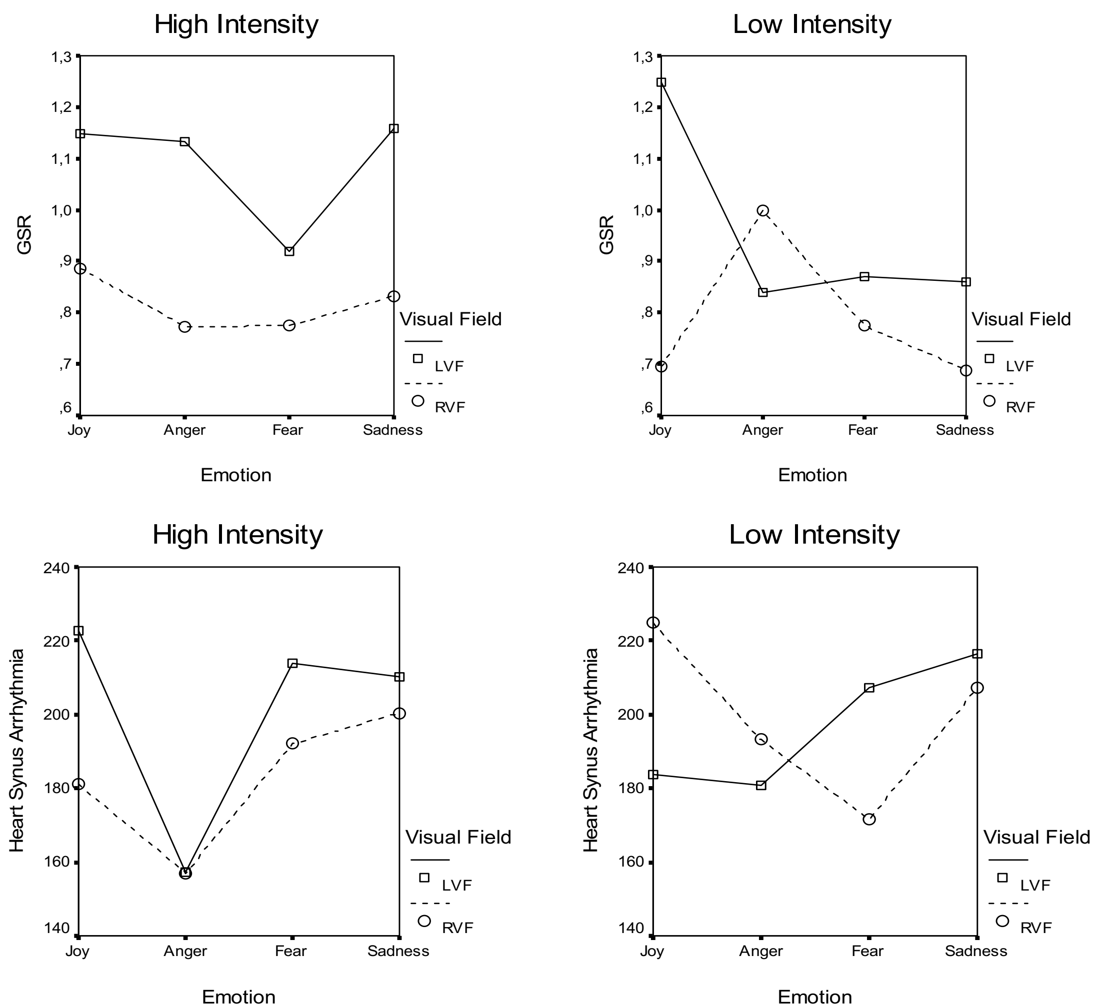


Figure 2. Factorial plots for GSR (μohms) and heart sinus arrhythmia (Emotion x Intensity x Visual Field)

Figure 2 graphically depicts the results obtained from autonomic measures (Galvanic Skin Response: upper row; Sinus Arrhythmia: bottom row). Despite differences associated with the levels of the intensity factor, both measures consistently reflect a general privilege of LVF over RVF, although more clearly so in the high intensity condition than in the low intensity one. Regarding the autonomic patterning of emotions, only Sinus Arrhythmia revealed a significant effect of the factor emotions for LVF ($F(3,42) = 3.999$; $p = .014$), specifically located between anger and sadness (pairwise comparisons with Bonferroni correction)

Turning now to CNS measures, significant main effects for emotion category as well as for intensity were found for both hemifields.

Figure 3 plots the results obtained at frontal and temporal scalp locations. Significant differences between emotions were observed at the high intensity condition ($p = .000$), with higher activation recorded by the left hemisphere EEG leads for joy and anger (positive values of asymmetry), and a reversed pattern of higher activation for fear and sadness revealed by the right hemisphere leads (negative values of asymmetry). Adding to the above reported findings on RT, this symmetrical pattern is thus once again in agreement with the approach-withdrawal view of lateralization.

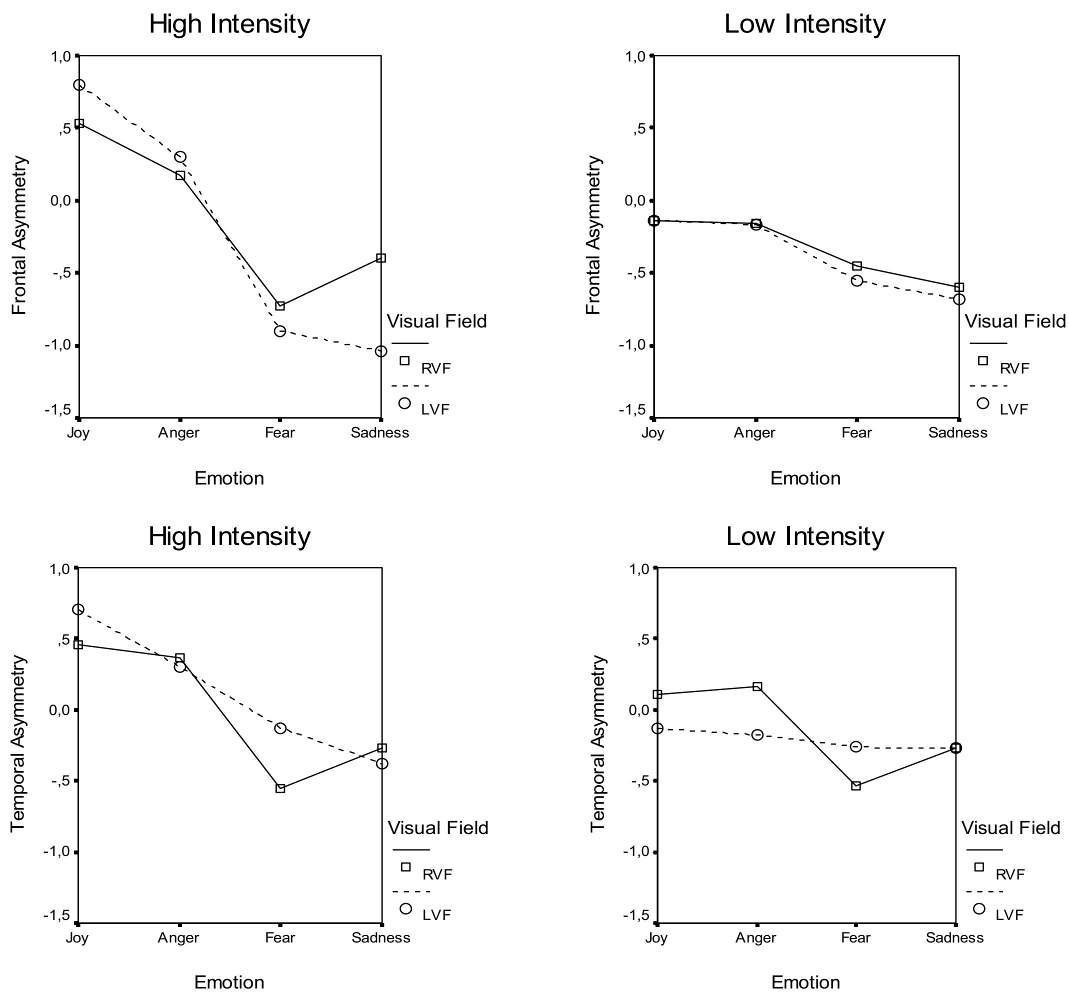


Figure 3. Factorial plots of the asymmetry scores of α-band power (Emotion x Intensity x Visual Field)

At low intensity levels, higher activation of the right hemisphere comparatively to the left one was generally seen (right panel: largely prevailing negative values of asymmetry). These results point toward an overall right hemisphere dominance, all across emotional category and affective valence, when low intensity levels are at stake.

Discussion

On the whole, outcomes lend support to the approach-withdrawal view of hemispheric lateralization whenever the high intensity condition is concerned, revealing a left hemisphere privilege for approach-related emotions (joy and anger) and a reversed left privilege for withdrawal-related emotions (fear and sadness). However, at the low intensity levels of emotion expression, overall right hemisphere dominance was instead observed, thus favoring the "right hemisphere theory" for low intensity conditions.

Regarding autonomic responses, right hemisphere dominance was largely the rule. This is in accordance with number of suggestions that autonomic regulation of emotions embodies a close preferential bond to the right hemisphere. On the other hand, our data couldn't provide but mixed evidence as regards emotion peripheral specificity, failing to demonstrate clear distinctive autonomic response patterns for emotion categories (although joy and anger may appear as the best candidates for that). In a sense, these comes as no surprise, given the limited number of autonomic channels assessed in the study, and the suggested existence of individual differences over preferred channels of autonomic expression (Vernet-Mauryn *et al*., 1999)

In light of these data, the behavioral approach-withdrawal orientation differentially entailed by distinct emotions appears as a major determinant of lateralization patterns, in accordance with Davidson's suggestions regarding the axial basis of prefrontal asymmetries. Furthermore, emotional intensity also emerged as second major factor, with the ability to shift or modulate lateralization patterns across different emotion categories.

A somewhat complex picture of lateralization effects, resting upon different dimensions for different dependent variables (e.g., CNS and ANS), comes out of our findings. Moreover, as a methodological stance, the need to systematically uncounfound intensity effects over the lateralization patterns, by routinely including intensity as a factor in the designs, is strongly warranted.

References

Borod, J.C., Cicero, B.A., Obler, L.K., Welkowitz, J., Erhan, H.M., Santschi, C., Grunwald, I.S., Agosti, R.M., Whalen, J.R. (1998). Right hemisphere emotional perception : Evidence across multiple channels. *Neuropsychology*, *12(3)*, 446-458.

Davidson, R.J. (2004). What does the prefrontal cortex «do» in affect : perspectives on frontal EEG asymmetry research. *Biological Psychology*, *67*, 219-233.

Grossman, P. (1992). Respiratory and cardiac rhythms as windows to central and autonomic biobehavioral regulation: selection of window frames, keeping the panes clean and viewing the neural topography. *Biological Psychology*, *34*, 131-161.

Harmon-Jones, E., Allen, J. (1998). Anger and frontal brain activity : EEG asymmetry consistent with approach motivation despite negative affective valence. *Journal of Personality and Social Psychology*, *74(5)*, 1310-1316.

Heller, W. (1993). Neuropsychological mechanisms of individual differences in emotion, personality and arousal. *Neuropsychology*, *7*, 476-489.

Reuter-Lorenz, P. & Davidson, R.J. (1981).Differential contributions of the cerebral hemispheres to the perception of happy and sad faces. *Neuropsychologia*, *19*, 609-613.

Sackeim, H.A., Gur, R.C., Saucy, M.C. (1978). Emotions are expressed more intensely on the left side of the face. *Science*, *202*, 434-436.

Tomarken, A.J., Davidson, R.J., Henriques, J.B. (1990). Resting frontal brain asymmetry predicts affective responses to films. *Journal of Personality and Social Psychology*, *59(5)*, 791-801.

Vernet-Maury, E., Alaoui-Ismaïli, O., Dittmar, A., Delhomme, G., & Chanel, J. (1999). Basic emotions induced by odorants : a new approach based on autonomic pattern results. *Journal of the Autonomic Nervous System*, *75*, 176-183.

Acknowledgment. This research was partially supported by grant POCTI/PSI/41235/2001 (FEDER/FCT).

REMEMBERED INSTRUCTIONS WITH SYMBOLIC AND PERCEPTUAL COMPARISONS

William M. Petrusic[1], Samuel Shaki[2], and Craig Leth-Steensen[1]
[1]Department of Psychology, Carleton University, Ottawa, K1S 5B6, Canada
[2]Department of Behavioral Sciences, College of Judea and Samaria, Ariel, 44837, Israel
bill_petrusic@carleton.ca; samuel_shaki@hotmail.com; craig_leth-steensen@carleton.ca

Abstract

Semantic congruity effects (SCEs) were obtained in each of two experiments, one with symbolic comparisons and the other with comparisons of visual extents. SCEs were reliably larger when the instructions indicating the direction of the comparison were represented by consonant-vowel-consonant (CVC) nonsense syllables, which had been associated with the conventional instructions in a preliminary learning phase of the experiment. The enhanced SCE with the CVC instructions is not readily admitted by any of the non-evidence accrual theories of the SCE (e.g., expectancy, semantic coding, and reference point). On the other hand, the general class of evidence accrual views of SCEs, such as those developed in Leth-Steensen and Marley (2000) and in Petrusic (1992), receive considerable empirical support.

Any complete theory of the process of comparing either perceptual or remembered stimuli must provide an explanation for the *semantic congruity effect* (SCE). The SCE is characterized by an interaction between the particular comparative instruction required and the location of the stimulus pair to be discriminated on the underlying continuum. For example, as in the landmark psychophysical experiments of Audley and Wallis (1964) that brought the SCE to the attention of the contemporary psychophysicists, the time to select the darker of two relatively dark lights is faster than the time to select the brighter. Conversely, selection of the brighter of two relatively bright lights is faster than the selection of the darker.

Although the main empirical properties of the SCE with perceptual and symbolic comparisons are now well established (e.g., Banks, 1977; Leth-Steensen & Marley, 2000; Petrusic, 1992), the empirical status of the available theories of the SCE remain, for the most part, to be firmly determined. The experiments reported here permit strong tests of the available theories of the SCE. These tests are based on the novel methodology in which participants first learn to associate consonant-vowel-consonant (CVC) nonsense syllables with comparative instructions and, subsequently, the magnitude of the SCE is gauged using both the conventional instructions and the symbolically represented CVC instructions.

The available classes of theories of the SCE can be viewed in terms of whether they are cast in the context of single instance, non-evidence accrual ideas, or embedded in terms of decision processing theories of the accrual of evidence. We specify the predictions for the various theoretical positions arising within each of these two broad frameworks in terms of the magnitudes of the SCE when instructions are presented directly as compared to when they are represented symbolically.

Additive stages, single-sample, non evidence accrual theories of the SCE

Semantic coding theory (Banks, 1977) asserts that the process of comparison involves discrete, strictly additive stages (i.e., coding of instructions, code activation of stimuli, code comparison, code translation if needed, and response selection). It predicts that though RTs will be lengthened with the CVC instruction due to slowed instructional access, the SCE should remain uninfluenced because the SCE occurs at a stage that follows the encoding of instructions. Specifically, the SCE occurs at the stimulus code translation stage, which necessarily follows the encoding of instructions. As well, the same instructional codes will be activated with the CVC instructions as with the conventional instructions. Consequently, the same code search and stimulus code translation processes will occur with the CVC instructions as with the conventional instructions. Thus, according to semantic coding theory, the SCE will be the same with CVC instructions as with the conventional instructions.

Expectancy theories (e.g., Marschark and Paivio 1981) posit that the instruction directs, much as in semantic priming, memory search for the relevant features of the to-be-discriminated stimulus pair. When the stimulus pair location is congruent with the instruction, the search process is semantically facilitated and when it is not, the search process must be redirected, slowing the comparison. In the present context, precisely the same expectancies are created with the CVC-based symbolic instructions as with the conventional instructions because precisely the same instructions are activated in each case. Consequently, the same search processes occur in the two cases and exactly the same SCE must occur with the two different forms of presentation of the instruction specifying the direction of the comparison.

Reference point theories (e.g., Jamieson & Petrusic, 1975) posit that stimuli are represented on an analogue continuum and presentation of an instruction activates an extreme point on the continuum, referred to as a reference point. Reference points are activated only upon encoding of instructions and it is expected that CVC-based instructions will slow speed of access to the relevant reference points. However, precisely the same reference points are activated with the CVC instructions as with the conventional instructions. Consequently, the SCE should be of precisely the same magnitude and form with the two different forms of instruction.

Evidence Accrual Theories

On the other hand, evidence accrual views of the SCE are clear in predicting that when primary decisional RTs are slowed due to the increased memory demands with remembered instructions, the SCE will also necessarily increase.

SCE as strategic bias. In the context of random walk-diffusion process models the SCE arises as a consequence of a dynamic, strategic adjustment of decisional criteria upon presentation of the instruction in the context of a particular stimulus pair (e.g., see Link 1992, pp. 172-178). Since precisely the same instructions are activated with the CVC and the conventional instructions, exactly the same biases occur with the two different ways of presenting the instructions for each trial. However, the drift of the accrual process in these models will be slowed given the increased memory requirements associated with distinguishing between the CVC instructions. Slower drift rates will enhance the time taken for the walk to reach the decision bounds resulting in a decisional latency effect that will be even more enhanced for cases, such as with semantically incongruent pairs, in which the boundaries have been set quite far from the starting point of the walk. Consequently, the SCE should be enhanced with the remembered, CVC, instructions.

Instructional pathway interference. Leth-Steensen and Marley (2000) have developed a connectionist-based evidence accrual model that posits the continuous accumulation of information about both the difference in stimulus magnitude and the end-point status of each stimulus item. This information is assumed to be accumulated simultaneously within two competing instructional pathways that are associated with both the relevant and irrelevant instructions, respectively. SCEs arise primarily on the assumption that the strengths of each instructional pathway, and, hence, the overall level of competition between them, are assumed to be dynamically modulated by the relative location of the stimulus items. Given that appropriate instructional pathway activation (or inhibition) also depends on the continuously available context provided by the instructions, and that such contextual information is less salient for the CVC instructions, the strength of the relevant instructional pathway will be weaker and, hence, instructional pathway interference will be greater with the CVC instructions than with the conventional instructions. Consequently, according to the Leth-Steensen and Marley model, SCE should be greater with the CVC instructions than with the conventional instructions.

Slowed evidence accrual. Petrusic (1992) has presented evidence that the SCE occurs at the level of each accrual in the accumulation of decisional evidence. Hence, any manipulation that further slows each accrual should result in an enhanced SCE. Because memory access to the relevant instruction will be slower with the CVC instruction than with the conventional instruction, if it is also the case that the CVC-instruction association must be activated on each evidence accrual, then the evidence accrual idea developed by Petrusic predicts a larger SCE with CVC instructions than with the conventional instructions.

EXPERIMENT 1

Method

Participants. Twenty Carleton University students participated in one 45-minute session to satisfy course requirements. All participants reported normal or corrected-to-normal vision.

Apparatus. Graphics production, presentation of instructions and stimuli, event sequencing and timing, and the recording of responses and RTs were controlled by a Pentium III computer running under SuperLab control. Stimuli and instructions were presented on a 17 inch (43 cm) ViewSonic video monitor with 800 by 600 pixel resolution. Responses were made using the buttons on an IBM-PC Mouse with the roller-ball disabled.

Stimuli. Twelve animals names, were used as the stimulus set. Six names were of relatively small animals (bee, rat, flea, crab, snail, and mouse) and the other six names were of relatively large animals (dog, pig, wolf, bear, horse, and whale). Three pairs of relatively small animals (bee-rat, flea-crab, snail-mouse) and three pairs of relatively large animals (dog-pig, wolf-bear, horse-whale) were created. The two words "Larger", "Smaller", and the eight nonsense syllables, GUF, BIX, NIQ, YOL, ZOE, KAG, LEG, and CEB, each with 30% association value, were used as instructions in a comparative judgment of size task.

Design and procedure. The session began with a learning phase in which the participants learned to associate each of the two instructional words ("Larger", "Smaller") with four of the eight CVCs. Following the learning phase, the participants were instructed that on each trial they would be presented with a pair of animal names, and either an instructional word or one of the CVCs they had learned to associate with the instruction words in the former phase. The participant's task was to press the mouse button on the side of the name of the larger (or smaller) animal in the pair of names, according the presented instruction. A given pair appeared in both spatial arrangements (i.e., each element in the pair, appeared once on the left and once on the right), and was shown with each of the two comparative instructions and with each of the eight CVCs. The instruction type (instructional word or CVC) occurred equally often with each pair. Four CVCs requiring selection of the smaller animal and four CVCs requiring selection of the larger animal accompanied each pair. As well, each of the conventional instructions accompanied each pair four times, thereby comprising a total of sixteen instructions. Each of the six stimulus pairs, two spatial arrangements of each pair, and 16 instructions was presented twice, for a total of 384 experimental trials, preceded by 16 practice trials.

Procedure. Participants were tested individually in a dimly lit room, seated approximately 80 cm from the centre of the screen. Participants initiated each trial by pressing both buttons of the mouse. Each trial then started with appearance of an instruction. After an additional 750 ms, the pair of animal names appeared while the comparative instruction remained on the screen. The stimuli and the instruction were response-terminated. The next trial began 1000 ms later. Participants were encouraged to respond quickly and accurately.

Results

The findings are presented in two main sections, the first presents response time (RT) analyses and the second focuses on error rates. For each participant, in all analyses, the dependent variables are either the mean RT for all responses or the mean percentage of errors in each cell of the design. In each analysis of variance (ANOVA) reported, the Huynh-Feldt epsilon adjustment of degrees of freedom was used. However, the degrees of freedom associated with each value of F are those defined by the design and the *MSE*s provided in the text are those given by the conventional degrees of freedom. The level of significance was set at .05 throughout.

The data of four participants were not used in the following analyses. Each of these participants was faster overall in responding with the CVC instructions than with the conventional instructions, therefore, denying the necessary condition for the tests of the alternative theories.

Response time analyses

Mean RTs with each instruction type for each stimulus pair in each instruction format condition are provided in Figure 1. An ANOVA confirmed a main effect of instruction format, $F(1, 15) = 31.01$, $MSE = 949581$. RTs in the conventional and CVC instruction conditions were 1425 and 1568 ms, respectively. The main effect of pair was also statistically reliable, $F(5, 75) = 12.51$, $MSE = 4914850$.

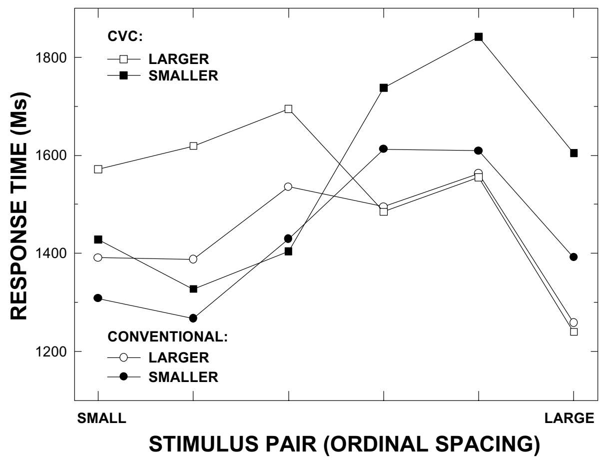


Figure 1. Mean response times with each instruction for each stimulus pair in the CVC and conventional instruction conditions in Experiment 1.

As is evident in the plots in Figure 1, the SCE occurs with both the conventional and the CVC instructions and the interaction between stimulus pair and instruction type is reliable, $F(5, 75) = 28.67$, $MSE = 28585$. Importantly, the three-way interaction involving instruction format condition, stimulus pair, and instruction type is significant, $F(5, 75) = 6.32$, $MSE = 25828$, affirming the fact that there is an enhanced SCE with the remembered, CVC, instructions.

Error Analyses

The correlation between mean RTs and mean error rate in each of the 24 cells defined by the factorial combination of instruction condition, instruction, and stimulus pair was positive, $r = .563$, $t(23) = 3.19$, $p < .004$, indicating that there was no speed-accuracy trade-off. Thus, generally, the effects evident with RTs also occurred with the error data. For example, more errors occurred with the CVC instructions (4.96%) than with the conventional instructions (3.95%) and error rate differed across pairs, although neither of these main effects was statistically reliable.

EXPERIMENT 2

With a view toward adding generality to the findings obtained in Experiment 1, strictly perceptual comparisons were required. Participants compared the lengths of simultaneously presented visual extents using both the conventional and the CVC instructions, as in Experiment 1.

Method

Participants. Twenty-one Carleton University students participated in two 50-min. sessions to satisfy course requirements. All participants reported normal or corrected-to-normal vision.

Apparatus. The apparatus of Experiment 1 was used.

Stimuli. Twelve horizontal lines were used as the stimulus set. Six lines were relatively short (10, 11, 20, 21, 50, and 51 pixels) and the other six were relatively long (147, 150, 200, 210, 250, and 252 pixels). Three pairs of relatively short lines (10-11, 20-21, 50-51) and three pairs of relatively long lines (147-150, 200-210, 250-252) were created. The three short stimulus pairs are defined, in terms of difficulty, by the ratios 1.10, 1.05, 1.02, respectively and the ratios are 1.02, 1.05, and 1.008 for the long pairs, respectively. All lines, drawn by Paintbrush software, were 1 mm wide and appeared in black on a white background. The pairs of lines appeared at the respective centres of the left and right hemi-fields on the monitor. The two words “Longer”, "Shorter" and the same eight nonsense syllables as in Experiment 1 were used as instructions in a comparative judgment of line length task.

Design and procedure. Each session began with a learning phase in which the participants learned to associate each of the two instructional words ("Longer", "Shorter") with four of the eight CVCs in the same manner as in Experiment 1. Following the learning phase, the participants were instructed that on each trial they would be presented with a pair of lines, one on the left and the other on the right, and either an instructional word or one of the CVCs they had learned to associate with the instruction words in the former phase. The participant's task was to press the mouse button on the side of the longer (or shorter) line, according the presented instruction. The remaining aspects of the design were the same as in Experiment 1 with the six pairs of animal names replaced by the six pairs of line lengths. Each 50-min session included one planned break, which ended with the participant's decision.

Results

Response time analyses

An ANOVA with the six pairs, the two instructions, and the two instruction conditions, showed the instructional manipulation to be effective. The main effect of instruction format was statistically reliable, $F(1, 20) = 12.02$, $MSE = 184740$. On average, RTs were 1336 ms and 1469 ms with the conventional and the CVC instructions, respectively. In addition, participants were 94 ms faster with the instruction to choose the shorter line, and this reverse lexical markedness effect was statistically reliable, $F(1, 20) = 18.46$, $MSE = 61307$. Comparisons varied systematically as function of both the stimulus ratio and the difference; the main effect of pair, $F(5, 100) = 3.35$, $MSE =79759$, was statistically reliable. For the short pairs, RTs were 1308, 1438, and 1432 ms for pairs with the ratios 1.10, 1.05, and 1.02, respectively. For the long pairs, the RTs were 1357, 1436, and 1444 ms for pairs with the ratios 1.05, 1.02, and 1.008, respectively.

The overall SCE was statistically reliable, $F(5, 100) = 7.46$, $MSE = 31925$. Importantly, the three-way interaction involving instruction format, pair, and instruction type was reliable, $F(5, 100) = 3.21$, $MSE = 28862$, confirming the enhancement of the SCE with the CVC instructions relative to the conventional instructions, as is evident from the plots provided in Figure 2.

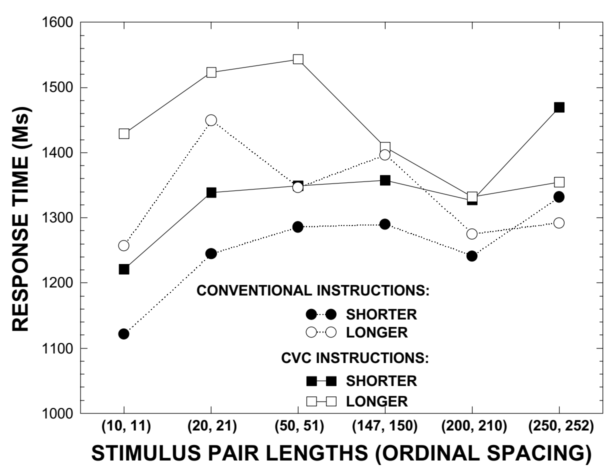


Figure 2. Mean response times with each instruction for each stimulus pair in the CVC and conventional instruction conditions in Experiment 2.

Error Analyses

As in Experiment 1, there was no evidence of a speed-accuracy trade-off, $r = 0.365$, $t(23) = 1.84$, $p < .08$ (two-tailed). Participants made fewer errors with the conventional instructions (31.54%) than with the CVC instructions (33.05%), mirroring the pattern obtained for RTs, $F(1, 20) = 6.62$, $MSE = 43.60$. In addition, as with RTs, the error percentage varied systematically with stimulus pair ratio and with difference when the ratio is held constant, $F(5, 100) = 73.07$, MSE = 119.23. For the short pairs, the error rates were 18.3%, 27.2%, and 41.0% for pairs with the ratios 1.10, 1.05, and 1.02, respectively. For the long pairs, error rates were 25.6%, 36.7%, and 44.9% for pairs with the ratios 1.05, 1.02, and 1.008, respectively. No other main effects or interactions attained statistical significance.

DISCUSSION AND CONCLUSIONS

The findings are clear and are not consistent with the single sample, expectancy, semantic coding, or reference point theories of the SCE. As indicated, each of these views of the SCE, while predicting increases in RTs with the CVC instructions is clear in predicting that the SCE should be the same with the CVC and the conventional instructions, contrary to the enhanced SCE that was observed with the CVC instructions. On the other hand, the general class of evidence accrual views of SCE, such as those developed in Leth-Steensen and Marley (2000) and in Petrusic (1992), receive considerable empirical support.

The present findings converge nicely with work reported in Shaki, Petrusic, and Leth-Steensen (2005). Shaki et al. showed, in each of two experiments requiring symbolic comparisons of animal size, that SCEs were enhanced when the instructions varied randomly from trial-to-trial as compared to when they were constant over a block of trials. They argued that their findings were not permitted according to any of the single sample (e.g., semantic coding, expectancy, or reference point) views of the SCE. Rather, precisely as in the present experiments, the enhanced SCE with the randomized instructions arose, in the context of the various evidence accrual theories, because the greater memory demands with the randomly varying instructions slowed the rate of accrual.

REFERENCES

Audley, R. J., & Wallis, C. P. (1964). Response instructions and the speed of relative judgments: I. Some experiments on brightness discrimination. *British Journal of Psychology*, **55**, 59-73.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgment. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101-159). New York: Academic Press.

Jamieson, D. G., & Petrusic, W. M. (1975). Relational judgments with remembered stimuli. *Perception & Psychophysics*, **18**, 373-378.

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review*, **107**, 62-100.

Link, S. W. (1992). *The wave theory of difference and similarity*. Hove and London: Erlbaum.

Marschark, M., & Paivio, A. (1981). Congruity and the perceptual comparison task. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 290-308.

Petrusic, W. M. (1992). Semantic congruity effects and theories of the comparison process. *Journal of Experimental Psychology: Human Perception and Performance*, **18**, 962-986.

Shaki, S., Petrusic, W.M., & Leth-Steensen, C. (2005). Effects of instruction presentation mode in comparative judgments. *Memory & Cognition*, In press.

ACKNOWLEDGEMENT

The work was supported by Natural Sciences and Engineering Research Council of Canada Discovery grants to William M. Petrusic, Craig Leth-Steensen, and Lise Paquet.

SEARCHING FOR THE INTERNAL CLOCK'S FUNDAMENTAL FREQUENCY

Marc Pouliot and Simon Grondin
Université Laval, Québec, Canada

Abstract

The goal of this study was to explore the frequential characteristics of the internal clock. Using the so-called Method of Dynamic Stimuli, developed to study tempo discrimination, the interval of indifference between 1.45 and 2.0 Hz was examined, making it possible to initially observe the behaviour of the internal clock in the vicinity of one of its fundamental frequencies of oscillation. The results of Experiment 1 revealed that (1) there was no difference in sensitivity to accelerations and decelerations at the center of the interval (1.7, 1.75 Hz), (2) decelerations were easier to detect with slower tempi, and (3) accelerations were easier to detect with faster tempi. In Experiment 2, where the range varied from 2.9 to 4 Hz (2nd harmonic), the same pattern of results was observed, indicating that the clock reacts with its second harmonic in the same way as it does with its first. However, in Experiment 3, where intervals ranging between the first and second harmonics (2 and 2.9 Hz) were tested, the results revealed that decelerations were easier to detect with slower tempi and that there was no sensitivity difference for accelerations or decelerations with higher tempi. This suggests that the internal clock was initialized at a higher frequency (3.4 Hz) than the frequency band used in the test and that this initialization could depend on the experimental context, i.e., in this case, the range of tempi used. Overall, the results in the different contexts suggest a fundamental frequency close to 1.7 Hz.

Many researchers have explored the characteristics of the internal clock (Church, 1984; Grondin 2001; Treisman 1963). Of particular interest in this regard is the search for the clock's referential frequency, which may lie within an indifference interval (Barnes & Jones, 2000; Fraisse, 1963; Jones & McAuley, 2005; Large & Jones, 1999; Treisman, 1963; Vos, van Assen & Franek, 1997). This interval is reported to show no difference in sensitivity for accelerations or decelerations.

If the internal clock has oscillatory characteristics and a reference frequency, it should react in the manner described by Vos & al. (1997). If a tempo is close to, but slower than a reference frequency, decelerations would be easier to detect since they do not follow the normal tendency to accelerate back to the reference frequency (see Figure 1). The opposite would be true for a tempo close to, but slightly faster than a reference frequency: in this case, accelerations would be easier to detect. If a tempo starts on the reference frequency, there would be no difference between the ease with which accelerations and decelerations are detected.

Based on the previous description, Figure 2 shows predicted response times in the vicinity of a reference frequency (left panel) and its first harmonic (right panel). If the starting frequency is lower than the reference frequency or its harmonic, the results should be similar, i.e., decelerations would be easier to detect; however, if the starting frequency is higher, accelerations would be easier to detect. Figure 3 illustrates the predicted response times for accelerations and decelerations for the first two harmonics of a reference frequency, i.e., the reference frequency itself (left panel) and its second harmonic (right panel). If the first harmonic is the reference, the starting frequency would always be higher, which should result in more sensitivity to accelerations, while if the second harmonic is the reference, the starting frequency would always be lower, which should result in more sensitivity to decelerations.

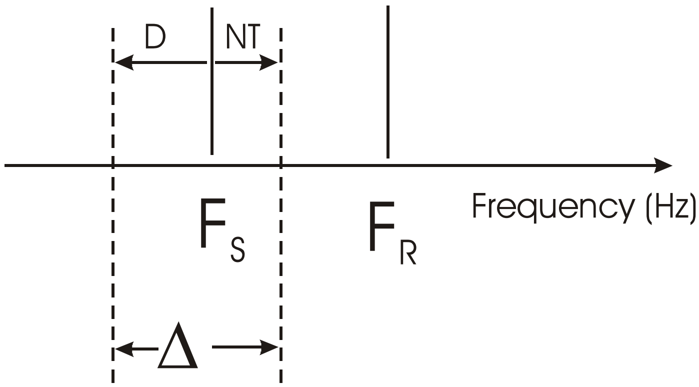


Figure 1. Oscillatory characteristics and presence of a reference frequency. The starting tempo F_S is close to, but slower than a reference frequency F_R, making decelerations (D) easier to detect, since they do not follow the normal tendency (NT) to accelerate back to the reference frequency (Δ greater).

While most discrimination methods are based on the relative number of correct and incorrect responses for determining difference thresholds, the so-called *Method of Dynamic Stimuli* (MSD) was developed for studying the referential frequency issue (Pouliot & Grondin, 2005). With this method, response times are used as the dependent variable. The method consists in presenting a series of reference tempi followed by gradually increasing or decreasing tempi. As soon as a change of tempo is perceived, the participant stops the stimulus and indicates if an acceleration or a deceleration was perceived. The MSD is argued to be a sensitive tool for studying slight tempo variations, and presents the advantage of revealing a given level of discrimination with only a few trials (Pouliot & Grondin, 2005).

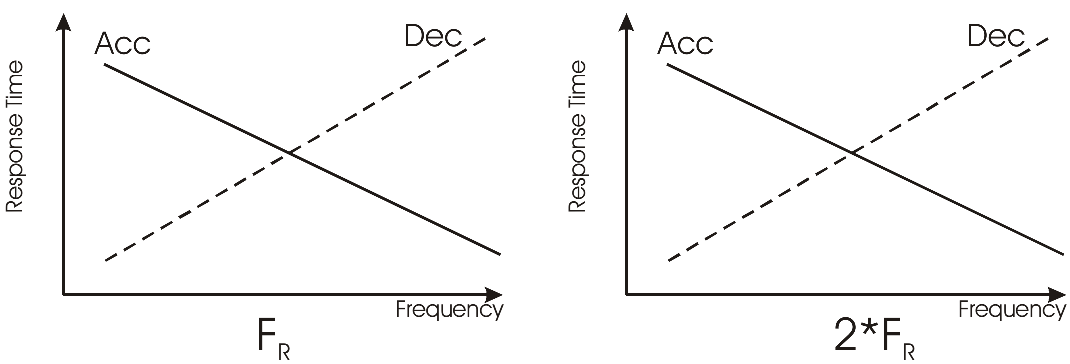


Figure 2. The left and right panels show, respectively, possible results in the vicinity of a reference frequency of an internal clock, and its second harmonic (Acc=Acceleration; Dec=Deceleration; F_R= Reference frequency; $2*F_R$=2nd harmonic of F_R).

The purpose of the following series of experiments was to explore the frequential characteristics of the internal clock. To that end, we analysed tempo discrimination performed with the MDS. Experiment 1 was carried out at a range between 1.45 and 2 Hz, which is often referred to as the indifference interval. In Experiment 2, tempo discrimination was performed at the second harmonic of the values used in Experiment 1, i.e., from 2.9 to 4 Hz. Finally, Experiment 3 was conducted at a range between those analyzed in the two preceding experiments (1.9 to 3.0 Hz).

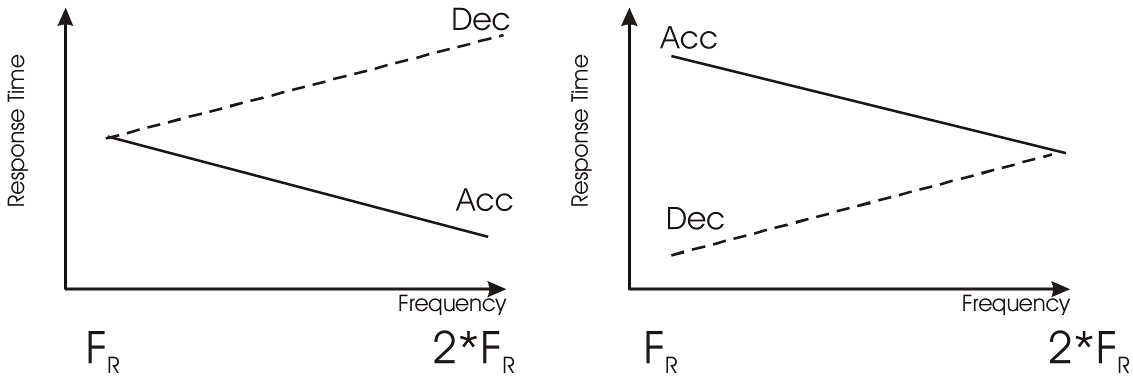


Figure 3. Possible results between two harmonics of a reference frequency. The left panel shows the results that might be obtained when the reference is set to the first harmonic, and the right panel, those that might be obtained if the reference is set to the second harmonic. (Acc=Acceleration; Dec=Deceleration; F_R= Reference frequency; $2*F_R=2^{nd}$ harmonic of F_R).

Experiment 1

Method

Fifteen Laval university students, 13 females (M = 21.5 years, SD = 1.5) and 2 males (M = 22.0 years, SD = 2.0), took part in the experiment. They received CAN$10 for their participation, and the experiment lasted approximately 60 minutes.

The experiment used starting tempi ranging from 1.45 to 2 Hz, investigated by steps of .05 Hz and in phase with an harmonic tonality (1000th) of 1450 to 2000 Hz. The stimuli had the following form before the tempo variations:

$\sin 2\pi F_1 T * \sin 2\pi F_2 T$ where F_1 = Initial Tempo; $F_2 = 1000 * F_1$

After 10 cycles of the initial tempo F_1, the stimuli had the following form:

$\sin 2\pi F_3 T * \sin 2\pi F_2 T$ where $F_3 = F_1 * (1+((\text{Sign})*\text{Slope}) * T)$; $F_2 = 1000 * F_1$

Five repetitions of each stimulus (5 repetitions x 12 tempi X 2 accelerations = 120 trials), as well as catch trials (12 tempi with no tempo variations), were presented randomly to the participants, for a total of 132 trials.

A repeated-measures factorial design was used on two independent variables: 12 starting Tempi (1.45 Hz to 2.00 Hz, with steps of .05 Hz) and 2 Signs (Acceleration or Deceleration). The dependent variable was the response time after the beginning of the tempo variations. The slope of variation was constant for the experiment and was set at 4%/sec during the tempo variations.

To ensure that participants understood the procedure, they were presented with a brief practice block with feedback, including 2 accelerations, 2 decelerations and 2 catch trials (without variation). The practice block was followed by the experimental session, which did not include feedback. The participants were instructed to be accurate, fast and consistent. Consistency was ensured by allowing them to take pauses between stimuli: delays between the stimuli were controlled by each participant.

Results

Participants committed several errors during the catch trials (19%), but very few in the trials with tempo variations (less than 1%). Figure 4 shows the mean response times for all participants. An ANOVA (Greenhouse-Geisser corrections were applied) revealed that there was no significant Sign effect (Acceleration vs. Deceleration), although the Tempo effect showed significant differences, $F(11,154) = 2.55$, $p < .05$. Most importantly, the Tempo*Sign interaction was significant, $F(11,154) = 5.81$, $p < .001$. It was in the range of 1.7 and 1.75 Hz, i.e., the range where it is most difficult to observe greater sensitivity to accelerations or decelerations.

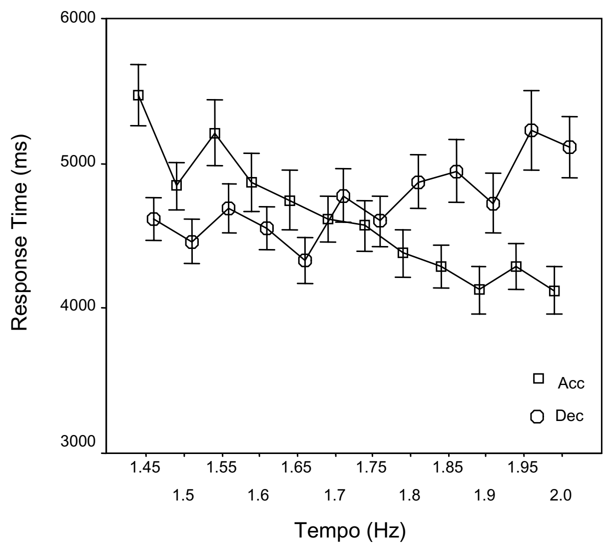


Figure 4. Response times to gradual accelerations and decelerations as a function of starting tempo (1.45 to 2.0 Hz). Bars represent standard errors.

Experiment 2

Method

Eighteen Laval university students, 15 females (M = 21.7 years, SD = 1.6) and 3 males (M = 23.5 years, SD = 3.3), took part in the experiment. They received CAN$10 for their participation, and the experiment lasted approximately 60 minutes.

Except in regard to the stimuli, which included the tempo range, the method used was similar to that employed in Experiment 1. Starting tempi ranged from 2.9 to 4.0 Hz, investigated by steps of .1 Hz and in phase with an harmonic tonality (256th) of 742.4 to 1024 Hz.

Results

Several errors were made during the catch trials (16.2%), but very few during the trials with tempo variations (1.5%). Figure 5 shows the mean response times for all participants. An ANOVA revealed that there was no significant Sign effect (Acceleration vs. Deceleration), although the Tempo effect showed significant differences, $F(11,187) = 3.93$, $p < .05$. Most importantly, the Tempo*Sign interaction was once again significant, $F(11,187) = 3.56$, $p < .01$.

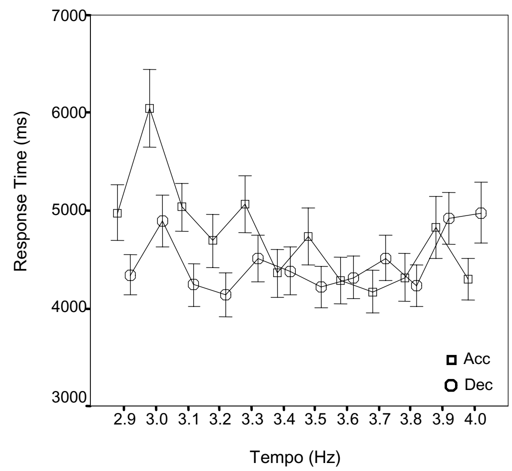


Figure 5. Response times to gradual accelerations and decelerations as a function of starting tempo (2.9 to 4.0 Hz). Bars represent standard errors.

Experiment 3

Method

Twenty Laval university students, 11 females (M = 25.7 years, SD = 6.7) and 9 males (M = 27.7 years, SD = 4.8), took part in the experiment. They received CAN$10 for their participation, and the experiment lasted about 60 minutes.

Except in regard to the stimuli, which included the tempo range, the method used was similar to that employed in Experiment 1. Starting tempi ranged from 1.9 to 3.0 Hz, investigated by steps of .1 Hz and in phase with an harmonic tonality (256th) of 486.4 to 768 Hz.

Results

Several errors were made during the catch trials (26.0%), but very few during the trials with tempo variations (0.9%). Figure 6 shows the mean response times for all participants. An ANOVA revealed significant Sign, $F(1,19) = 11.80$, $p < .01$, and Tempo effects, $F(11,209) = 7.22$, $p < .001$. Interestingly, the Tempo*Sign interaction was also significant, $F(11,209) = 3.41$, $p < .05$.

General Discussion

The Tempo*Sign interaction was significant for all three experiments, which is consistent with the predicted results. In experiments 1 and 2, the Sign effect was not significant, which indicates that the response times for accelerations and decelerations were equal, another finding that is consistent with the predictions described above: there should be no difference between accelerations and decelerations when a tempo is centered on a reference frequency. The results of Experiment 1 were closer to the initial predictions than those of Experiment 2, in that there was only a small range where response times to accelerations and decelerations overlapped.

Although the results of Experiment 2 were less conclusive than those of Experiment 1, the expected x-shaped interaction was observed. The results of Experiment 3, where the starting tempi ranged between the reference frequency and its harmonic, exhibit a pattern consistent with the one illustrated on the right panel of Figure 3. It seems that the internal clock may have set itself to the second harmonic, thereby demonstrating adaptability.

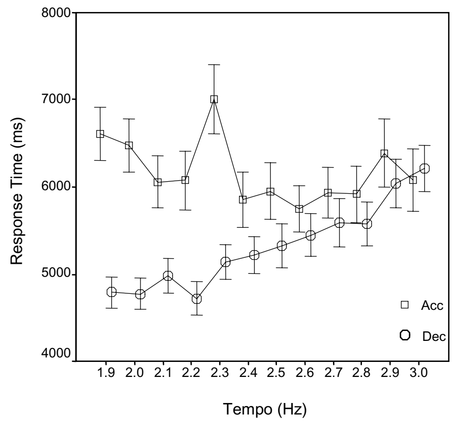


Figure 6. Response times to gradual accelerations and decelerations as a function of starting tempo (1.9 to 3.0 Hz). Bars represent standard errors.

The results of these three experiments also confirm the hypotheses for oscillatory characteristics. In a way, the results bridge two models: the beat-based model with its oscillatory characteristics (Treisman, 1963; Vos & al, 1997) and the entrainment model with its adaptive characteristics (Jones & McAuley, 2005; Large & Jones, 1999). Beat-based models are often described as non-adaptive, and entrainment models are not linked to fixed oscillatory characteristics. In this study, however, an "adaptive beat-based" model predicts the results appropriately. The present experiments illustrate the oscillatory characteristics of the internal clock and its adaptability. The results suggest that an internal clock's reference frequency, or fundamental frequency, is close to 1.7 Hz.

References

Church, R. M. (1984). Properties of the Internal Clock. *Annals of the New York Academy of Sciences*, 423, 566-582.

Fraisse, P. (1963). Perception et estimation du temps (Perception and estimation of time). In P. Fraisse & J. Piaget (eds.), *Traité de psychologie expérimentale*: Vol 6, La perception (pp. 59-95). Paris: Presses Universitaires de France.

Grondin, S. (2001). From Physical Time to the First and Second Moments of Psychological Time. *Psychological Bulletin*, 127, 22-44.

Large, E. W. & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119-159.

Jones, M. R. & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Perception and Psychophysics*, 67, 398-417.

Pouliot, M. & Grondin, S. (2005). A response time approach for estimating sensitivity to auditory tempo changes. *Music Perception*, 22, 389-399.

Vos, P. G., Van Assen, M. & Franek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research*, 59, 240-247.

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the ''internal clock.'' *Psychological Monographs*, 77, 1–31.

SENSITIVITY TO TEMPO CHANGES: A PROCEDURAL EFFECT

Marc Pouliot, Marilyn Plourde, Simon Grondin and Normand Teasdale
Université Laval, Québec, Canada

ABSTRACT

The purpose of this experiment was to test the potential effect of a repeated-measures design in the study of internal timekeeping processes. In a series of brief auditory signals, intervals between signals were slowly increased or decreased, and participants were asked to detect the direction of the temporal changes (acceleration or deceleration). Different acceleration or deceleration conditions were tested, with participants being randomly assigned to begin the experiment in the difficult (slow changes) or the easy (rapid changes) condition. The results revealed an asymmetrical effect of procedure: when participants began with the most difficult condition, their performance in the easier condition was not affected, but when they began with the easier condition, their performance in the more difficult condition declined .

The goal of this experiment was to investigate the potential procedural effect of a repeated-measures design using the Dynamic Stimuli Method (Pouliot & Grondin, 2005), whereby a starting tempo is gradually accelerated or decelerated. This method is used for studying relative sensitivity to acceleration and deceleration. We used a starting tempo of 102 beats per minute (bpm), which is often referred to as the center of a so-called indifference interval (Fraisse, 1963; Large & Jones, 1999; Vos, van Assen & Franek, 1997).

In the present experiment, we addressed the question of the order of presentation of two different slopes of tempo variations. Poulton and Freeman (1966) suggested that plans with repeated measures can reduce or exaggerate the effect between two experimental conditions. The present experiment did not initially show the expected interaction, when repeated-measures analysis was used. However, the expected interaction was observed when the data were analysed in the way Poulton and Freeman proposed.

Method

Participants

Twenty volunteer students from Laval University, 10 women (age, mean: 21.7 years, SD: 3.3) and 10 men (age, mean: 22.7 years, SD: 5.4), took part in the experiment. They received CAN$5.00 for their participation, and the experiment lasted approximately 30 minutes.

Apparatus and stimuli

A Pentium IV computer was used with a program developed with E-Prime. The program presented the instructions and audio stimuli to the participants and recorded information related to the stimuli, answers and response times. All stimuli were presented via headphones (Sony Mdr-6000) connected to the headphone output of a soundcard (Audigy 2 by Creative Labs) with a latency of 0.2 ms.

Participants responded by pressing buttons (Stop, + (acceleration), - (deceleration), Next) on an SR-box (Serial Response Box - E Prime) providing a resolution of 1 ms.

Three auditory stimuli (92-ms pulses at 521.05 Hz, 1042.5 Hz and 2085 Hz) were presented, with a starting tempo of 102 bpm. Ten pulses producing the starting tempo were presented and then followed by a tempo acceleration or deceleration (referred to as sign). The slope of these tempo changes was either 1%/sec (difficult to detect: S1) or 3%/sec (easier to detect: S3). All participants completed 60 trials (5 repetitions of 3 frequencies x 2 slopes x 2 signs).

Procedure

All 20 participants completed the two experimental blocks (1%/sec and 3%/sec). Half were randomly assigned to start with the 1%/sec-slope condition (Group 1: G1), while the other participants (Group 2: G2) began with the 3%/sec-slope. The stimuli (5 x 3 x 2) were presented in the same random order within each half of the experiment. The mean of the five trials was used as final data; it included only good responses and the dependent variable was the corresponding response time. Overall, we recorded only 62 wrong responses and the number of wrong responses was not affected by the starting condition.

Results

Several different analyses can be conducted with the present experimental procedures to examine the potential effect of using a repeated-measures design. The following statistical analyses will be presented: (Case 1) a conventional repeated-measures design analysis (randomized block factorial design); (Case 2) a completely randomized factorial design analysis based on the block or the starting slope (Group); and (Case 3) a completely randomized factorial design analysis using the starting group as an independent variable to explore the differences between groups.

In Case 1, or the factorial repeated-measures design analysis, all of the data pertaining to all of the participants were considered in the analysis (G1S1, G1S3, G2S3 and G2S1: see Figure 1). In Case 2, the completely randomized factorial design, only half of the data were used for the final analysis: either the data from the first block (G1S1 and G2S3) or the data from the second block (G2S1 and G1S3) (see Figure 1, left panel).Finally, in Case 3, or the completely randomized factorial design based on the starting group, all of the data ([G1S1+G1S3] and [G2S3+G2S1]) were used (see Figure 1, right panel).

The following factors were analyzed in the repeated-measures design (Case 1): 3 frequencies X 2 signs X 2 slopes. The ANOVA revealed a main effect for Slope only, $F\ (1,19) = 95.15$, $p < .001$., All other effects were not significant. It should be noted, however, that the interaction of interest here, i.e., Slope*Sign, was marginally significant, $F(1,19) = 3.87$, $p = .06$. Figure 2 shows the Response times according to Slope and Sign for all the data ($n = 20$). The Greenhouse-Geisser correction was used.

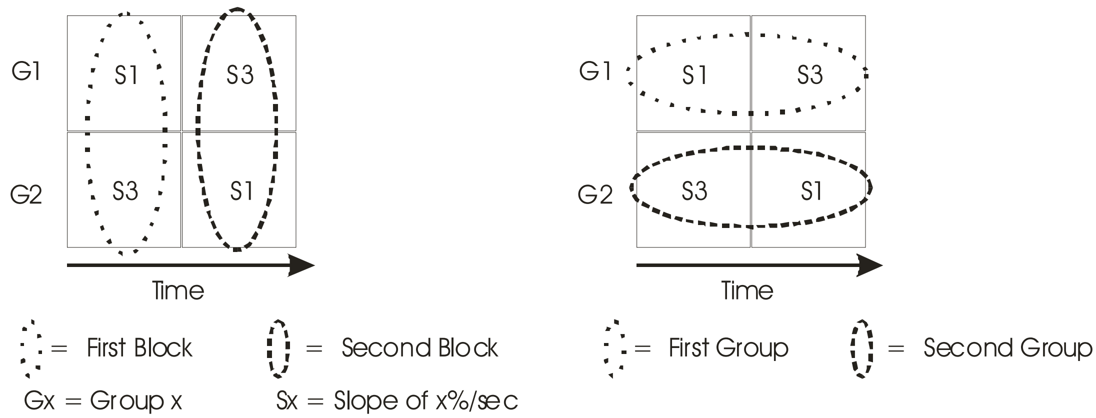


Figure 1. Diagram of possible ways of analyzing the data from the present experiment, using half of the data to compare the first and second blocks (Case 2) (left panel) or the first and second groups (Contrasts of Case 3) (right panel).

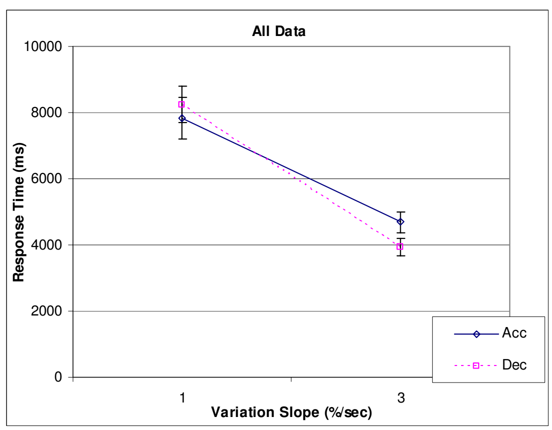


Figure 2. Mean response time for accelerations and decelerations as a function of tempo slope. Bars represent standard errors.

In Case 2, the graphical analyses of the data point to the presence of an interaction between the Slope and Sign factors (see Figure 3). ANOVAs were conducted to examine the relation between the raw data without removing part of the error related to the repeated-measures calculations.

In the random factorial design using the first or second blocks (Case 2), a 3 X 2 X 2 ANOVA (Frequency, Sign, Slope) revealed a significant main effect for Slope for the first block (G1S1 and G2S3), with $F(1,563) = 203.69$, $p < .001$. For the second block (G1S3 and G2S1), the ANOVA also revealed a significant main effect for Slope, with $F(1,573) = 223.37$, $p < .001$, and a significant Slope*Sign interaction, $F(1,573) = 15.46$, $p < .001$. Figure 3 shows the Sign effects (Acceleration vs Deceleration) for the various tempo variation slopes using available data ($n = 10$).

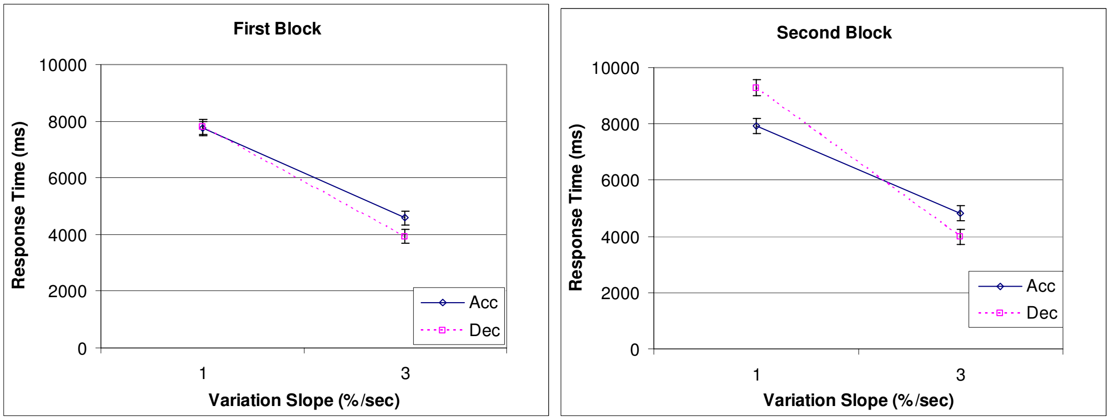


Figure 3. Slope*Sign interaction for the blocks. The first block (G1S1 and G2S3) is shown in the left panel, and the second block (G1S3 and G2S1) in the right. Bars represent standard errors.

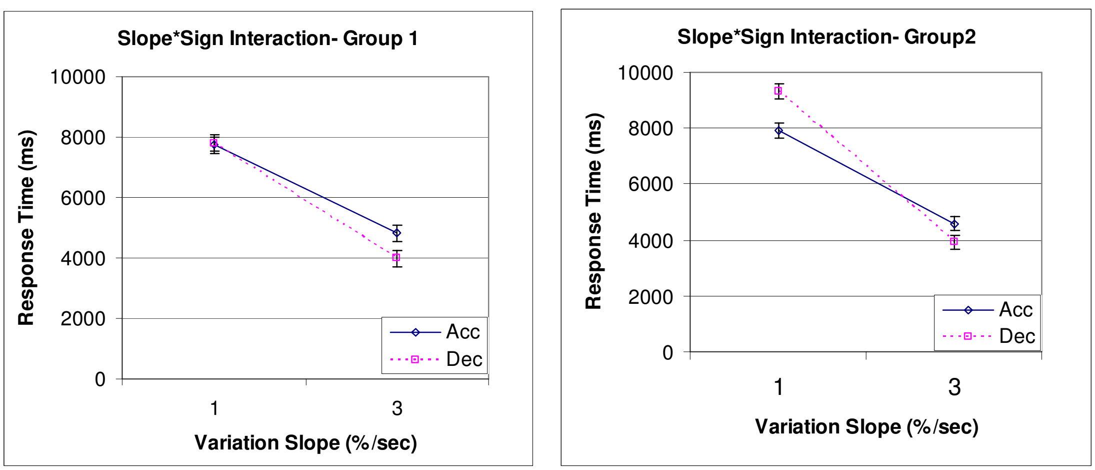


Figure 4. Slope*Sign interaction for the groups. The Group-1 condition (G1S1 and G1S3) is shown in the left panel and the Group-2 condition (G2S1 and G2S3) in the right. Bars represent standard errors.

Finally, a completely random factorial design (Case 3) using the starting group as an independent variable was employed. The purpose of this unorthodox way of examining the data was to identify a potential asymmetrical transfer. A 3x2x2x2 ANOVA (3 frequencies X 2 signs X 2 slopes X 2 groups (group beginning with 1%/sec or 3%sec)), revealed a significant main effect for Slope, $F(1,1137) = 425.16$, $p < .001$, with interactions of Group*Slope, $F(1,1137) = 7.13$, $p < .01$, Group*Sign, $F(1,1137) = 3.98$, $p < .05$, Slope*Sign, $F(1,1137) = 15.35$, $p < .001$ and Frequency*Sign, $F(2,1137) = 3.38$, $p < .05$. The Frequency*Sign interaction will not be discussed. Figure 4 shows the data for Slope*Sign interactions according to group. In comparisons of the groups, the ANOVAs (3x2x2 - Frequency X Sign X Slope) revealed a significant Slope effect for Group 1 (beginning with a Slope of 1%/sec) and for Group 2 (beginning with a Slope of 3%/sec) with $F(1,566) = 170.21$, $p < .001$ and $F(1,570) = 257.61$, $p < .001$ respectively. The Slope*Sign interaction was only present for Group 2 with $F(1,570) = 14.12$, $p < .001$. In comparisons of the slopes, ANOVAs with a Slope of 1%/sec revealed a significant effect for Group with $F(1,550) = 5.76$, $p < .05$ and for Sign with $F(1,550) = 4.15$, $p < .05$, while a Slope of 3%/sec revealed a significant effect only for Sign with $F(1,586) =$

25.39, $p < .001$. Figure 5 illustrates the interaction between slopes and groups, which highlights a Group difference for the Slope of 1%/sec (S1).

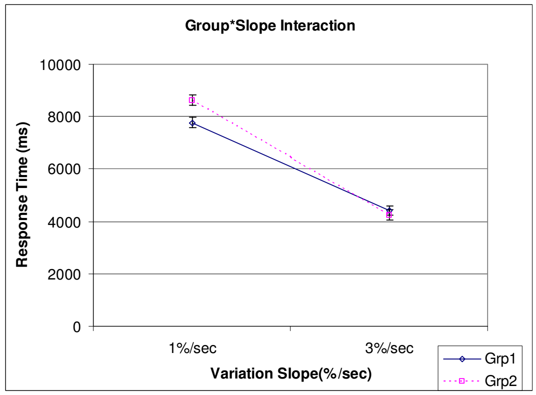


Figure 5. Response time for groups as a function of tempo variation slopes. Bars represent standard errors.

Discussion

The main reason for presenting these different analyses stems from a controversy concerning repeated-measures experimental designs. Poulton and Freeman (1966) suggested that plans with repeated measures can generally reduce, or sometimes exaggerate, the effect between two conditions.

The first experimental design distinguishes between two ways of grouping: Group 1-S1S3 (S1 condition performed first, followed by the S3 condition) and Group 2-S3S1 (S3 condition performed first, followed by the S1 condition). If no procedural effect occurs, there is no transfer between the conditions. However, if an equivalent practice effect is generated for the two groups by beginning the experiment with either the S1 or the S3 condition, a symmetrical transfer is observed, which does not harm the end result. A one-way asymmetrical transfer occurs if, for example, the S1 condition remains unchanged from one group to another, but performance on S3 improves or worsens after carrying out S1. A two-way asymmetrical transfer occurs if, for example, the S3 condition is better after the S1 condition and the S1 condition is worse after the S3 condition (or vice versa). In the case of asymmetrical transfers, the differences between the groups and the conditions are no longer equivalent. These inequalities suggest that results showing asymmetrical transfers should be reconsidered.

Poulton and Freeman (1966) recommended making an inter-subjects analysis of the first blocks before proceeding with the repeated-measures analysis and stressed that the criteria of nonadditivity, reflected by the tests of Tukey or Mauchly, can indicate the presence of an asymmetrical transfer. In the present case, even when the Greenhouse-Geisser correction was used, the Sign*Slope interaction was not significant when either the repeated-measures analysis or the inter-subjects analysis of the first blocks was employed. However, it was significant when the analysis of the second blocks was used. Under these conditions, the unconventional analysis of the groups also suggests that there was a Slope*Sign interaction for Group 2. When the data of the first and second blocks (see Figure 3) are compared with those of the groups (see Figure 4), the following hypothesis can be proposed. The data relating to blocks and groups seem to be equivalent for G1S1-G2S3 and G1S1-G1S3 and for G1S3-G2S1 and G2S1-G2S3 (see Figure 6).

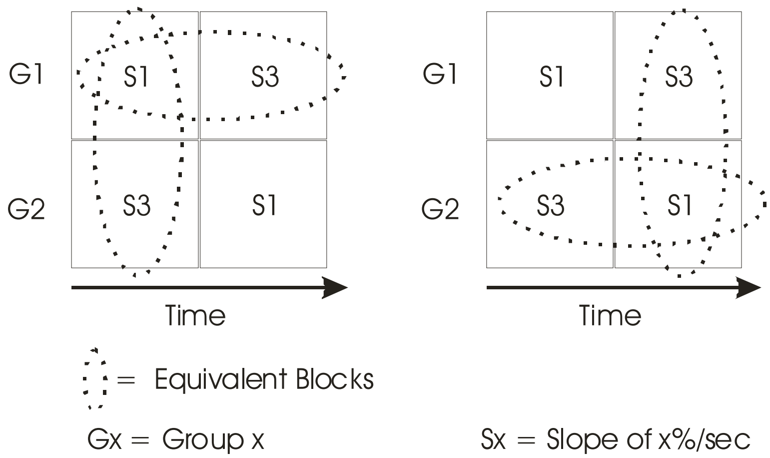


Figure 6. Equivalent blocks and groups. Data are equivalent for G1S1-G2S3 and G1S1-G1S3 as well as for G1S3-G2S1 and G2S1-G2S3. A difference was observed between the S1 conditions, while the S3 conditions were identical.

A difference can be observed for the S1 condition (most difficult) but not for the S3 condition (least difficult), and these findings are supported by the above-mentioned ANOVAs. In the current experiment, starting with the easier condition led to a decrease in performance in the most difficult condition. Therefore, using a repeated-measures design can hide the Slope*Sign interaction that was present for Group 2 but not for Group 1. Moreover, the current results and analyses strongly suggest that caution should be exercised in rejecting an effect or an interaction using a repeated-measures design. Since the Dynamic Stimuli Method does not require a lot of stimuli to provide sufficient statistical power, experiments can easily be adapted to avoid the potential procedural effect of the starting slope.

References

Fraisse, P. (1963). Perception et estimation du temps (Perception and estimation of time). In P. Fraisse & J. Piaget (Eds.), *Traité de psychologie expérimentale*: Vol 6, La perception (pp. 59-95). Paris: Presses Universitaires de France.

Large, E. W. & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106, 119-159.

Pouliot, M. & Grondin, S. (2005). A response time approach for estimating sensitivity to auditory tempo changes. *Music Perception*, 22, 389-399.

Poulton, E. C. & Freeman, P. R. (1966). Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin*, 66, 1-8

Vos, P. G., Van Assen, M. & Franek, M. (1997). Perceived tempo change is dependent on base tempo and direction of change: Evidence for a generalized version of Schulze's (1978) internal beat model. *Psychological Research*, 59, 240-247.

HEDONIC CONTRAST IN EXPERTS AND NON-EXPERT: CATEGORIZATION REVISTED

Lauren Rota and Debra A. Zellner
Department of Psychology, Montclair State University, Montclair, NJ 07043, USA
rotal@mail.montclair.edu and zellnerd@mail.montclair.edu

Abstract

Test stimuli are rated as less "good" following very good context stimuli than when presented alone. This diminution in rating is called hedonic contrast. Contrast is attenuated if context and test stimuli are perceived as being in different categories. Because experts use as their basic level category the subordinate level of novices they will categorize when novices do not. So, in the following studies both experts and novices showed hedonic contrast when attractive context orchids preceded more neutral test orchids. However, only the novices showed hedonic contrast when attractive context iris preceded the test orchids. Novices viewed the iris and orchids as "flowers" and therefore members of the same category, resulting in contrast. Experts viewed the iris and orchids as being in different categories and therefore hedonic contrast was eliminated.

Hedonic contrast was first described by Fechner in 1898 according to Beebe-Center (1932/1965, p.222). Fechner stated his general principle of hedonic contrast as follows: "That which gives pleasure gives more pleasure the more it enters into contrast with sources of displeasure or of lesser pleasure; and a corresponding proposition holds for that which gives displeasure: (p. 232, attributed to Fechner, 1898). According to Beebe-Center, Fechner believed that hedonic contrast would only occur if certain conditions were fulfilled, including that "the two factors had to bear a certain resemblance to each other" (pp. 222-223). That is, the stimulus to be judged (hereafter test stimulus) and the context stimuli had to resemble one another or be from the same category of stimuli.

We (Dolese, Zellner, Vasserman, & Parker, 2005; Zellner, Rohm, Bassetti, & Parker, 2003) recently demonstrated this phenomenon in laboratory studies where subjects were instructed to consider the context and test stimuli either to be in the same category or in different categories. Contrast was attenuated when subjects were instructed to view the context and test stimuli as being in different categories. So, for example (Zellner et al., 2003), a group of pictures of North American birds that followed a group of pictures of brightly colored Tropical birds were judged as less attractive by subjects who were told that all of the stimuli were "birds" than by subjects who were told that the first set were "Tropical birds" and the second set were "North American birds". We (Zellner, Kern, & Parker, 2002) found similar effects on hedonic evaluation among subjects who said they viewed context and test stimuli (e.g., gourmet and canned coffees) as belonging to the same category (hedonic contrast was found) or different categories of stimuli (attenuation of hedonic contrast was found). Thus, viewing context and test stimuli as being from different categories attenuates the hedonic contrast seen when the two sets of stimuli are viewed as being from the same category.

People categorize things at different levels. The levels are usually referred to as superordinate, basic, and subordinate. For most people a basic level category is one in which objects are recognized most rapidly and at which adults spontaneously identify objects (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). What is considered a basic category for an expert is different from that for a novice. Studies (Johnson & Mervis, 1997; Tanaka & Taylor, 1991) have found that what is a novice's subordinate level is an expert's basic level.

The following experiments investigate the effect of naturally-occurring categories on hedonic contrast by testing some people who are flower novices and others who are flower experts. Since the basic taxonomic level for experts is the subordinate level for novices we expect the experts to separate types of flowers (in this case iris and orchids) into two categories whereas novices should perceive them as being in one category (e.g., flower). This difference in level of categorization should result in hedonic contrast in novices when attractive iris precede hedonically neutral test orchids but less or no contrast in experts.

Experiment 1 – Flower Novices

This experiment investigates the effect of natural categorization on hedonic contrast in flower novices. Since for these subjects, both orchids and iris are in the same category, "flowers", and are therefore compared, hedonic contrast should result when attractive versions of either flower precede more hedonically neutral test orchids.

Method

Participants

Forty-three volunteers (undergraduates, graduate students, staff, and faculty) from Montclair State University served as subjects. All were flower novices; none of them had knowledge of or experience with flowers that would qualify them as flower experts. Thirty four of the participants were female and nine were male. The participants' mean age was 24.3 years.

Materials

Colored photographs of flowers cut out of books were pasted on 7 x 11 inch white posterboard. One set of seven pictures was of iris (context iris) rated as the most attractive out of a larger group in preliminary testing. A second set of seven pictures were of orchids (context orchids) rated as the most attractive out of a larger group in preliminary testing. A third set of six pictures was of orchids (test orchids) rated as neutral or only slightly attractive out of a larger group in preliminary testing.

Procedure

Subjects were randomly assigned to one of three groups: IO, OO, and C (explanations follow). All subjects were told that they would be rating pictures of flowers.

People in Group IO (Iris-Orchid, n=13) were shown a set of seven pictures of attractive iris (context iris) followed by a set of six pictures of only slightly attractive orchids (test orchids).

People in Group OO (Orchid-Orchid, n=15) were shown a set of seven pictures of attractive orchids (context orchids) followed by the set of six pictures of only slightly attractive orchids (test orchids).

People in Group C (Control, n=15) were only shown the set of six pictures of only slightly attractive orchids (test orchids).

Thus, people in Groups IO and OO rated a set of attractive context flowers prior to rating the set of test orchids. However, Group C only rated the test orchids.

All participants were tested individually. The participants in all groups rated the attractiveness of each flower on a 201-point bipolar hedonic scale on which +100 represented "the most attractive flower imaginable", -100 represented "the most unattractive flower imaginable," and 0 indicated a flower that was "neither attractive nor unattractive." They inspected one picture at a time (stimulus sequences randomized within sets) and announced their ratings.

Following the ratings of the flowers subjects filled out a questionnaire asking them their gender, age, and occupation. In order to asses their knowledge of flowers they were also asked to indicate how many types of flowers they saw, what kinds of flowers they saw (e.g., were they all tulips), and if the flowers they saw were arranged in a particular order. In addition, they were asked if they belonged to a garden club, had taken classes in flower arranging, or ever worked in a flower shop. They also were asked if they considered themselves to be flower experts.

Results

For each subject in Group IO we calculated the average rating for the context iris (M=53.7, SD=16.5) and for each subject in Group OO we calculated the average rating for the context orchids (M=35.8, SD=28.6). Although the context iris were judged somewhat more attractive than the context orchids they were not significantly so, Mann-Whitney (n_1=13, n_2=15) U_1=137, U_2=58, p>.06).

For each subject in each group we calculated the average rating assigned to flowers in the test orchid set. A Kruskal-Wallis test revealed that the attractiveness ratings of the test orchids differed among the groups, H(2)=10.4, p<.006. Post-hoc contrasts using alpha levels of .05 (Hollander & Wolfe, 1973) showed that both Groups IO (M=4.7, SD=32.8) and OO (M=0.3, SD=28.2) rated the test orchids as less attractive than did Group C (M=36.4, SD=27.4). There was no significant difference in the ratings of the test orchids by Groups IO and OO. See Table 1.

Table 1
Mean Ratings and Standard Deviations of Context and Test Flowers for the Three Groups of Novices

Group	Context Flowers (Iris or Orchids)	Test Orchids
IO	53.7 (16.5)	4.7 (32.8)
OO	35.8 (28.6)	0.3 (28.2)
C	----	36.4 (27.4)

Discussion

As expected, we found hedonic contrast in Group OO who rated attractive context orchids before the test orchids. These subjects rated the test orchids very close to hedonic neutrality (0.3) compared to the control subjects, Group C, who rated them as fairly attractive (36.4). We also found hedonic contrast in Group IO who also rated the test orchids very close to hedonic neutrality (4.7). Although the iris and orchids are in different subordinate categories, the novices in this study did not see them that way. Because their basic category was "flowers" they viewed the context iris and test orchids as being in the same category. This resulted in hedonic contrast in both Group IO and Group OO where attractive context flowers preceded the test orchids.

Experiment 2 – Flower Experts

The categorization of flowers into subordinate categories is much more likely to occur among people who are experts. If experts use as their basic taxonomic level in their area of expertise what is the subordinate level of novices (Johnson & Mervis, 1997; Tanaka & Taylor, 1991) we would expect that experts would show contrast when attractive orchids precede the test orchids but not when attractive iris precede the test orchids, (i.e., the hedonic categorization effect).

Method

Participants

Thirty volunteers from flower shops, garden clubs and other horticultural organizations served as subjects. All subjects had training and experience working with flowers. Twenty-one of the participants were female and nine were male. The participants' mean age was 43.9 years.

Materials

The materials used were the same materials used in Experiment 1.

Procedure

The procedure was the same as described in Experiment 1. As in Experiment 1, participants were randomly assigned to the three groups. All groups had 10 participants.

Results

For each subject in Group IO we calculated the average rating for the context iris (M=61.5, SD=26.8) and for each subject in Group OO we calculated the average rating for the context orchids (M=46.1, SD=14.0). Although the context iris were judged somewhat more attractive than the context orchids they were not significantly so, Mann-Whitney (n_1=10, n_2=10) U_1=35, U_2=65, p>.25).

For each subject in each group we calculated the average rating assigned to flowers in the test orchid set. A Kruskal-Wallis test revealed that the attractiveness ratings of the test orchids differed significantly among the groups, H(2)=8.8, p<.02. Post-hoc contrasts using alpha levels of .05 (Hollander & Wolfe, 1973) showed that Group OO (M=14.0, SD=39.2) rated the test orchids as less attractive than did Group C (M=52.8, SD=18.8). There was no significant difference in the ratings of the test orchids by Groups IO (M=59.2, SD=25.1) and C. There was a significant difference in the ratings of the test orchids by Groups OO and IO. See Table 2.

Table 2
Mean Ratings and Standard Deviations of Context and Test Flowers for the Three Groups of Experts

Group	Context Flowers (Iris or Orchids)	Test Orchids
IO	61.5 (26.8)	59.2 (25.1)
OO	46.1 (14.0)	14.0 (39.2)
C	----	52.8 (18.8)

Discussion

Not surprisingly, the flower experts rated the test orchids as more attractive than did the novices. As with the novices, we found hedonic contrast in Group OO who rated attractive context orchids before the test orchids. The experts in Group OO rated the test orchids close to hedonic neutrality (14.0) compared to the control subjects, Group C, who rated them as fairly attractive (52.8). Unlike the novices, the experts in Group IO did not show hedonic contrast. Because their basic level category consisted of "iris" and "orchids" experts viewing the attractive iris before the test orchids did not compare them. Having the context and test flowers be from different categories resulted in elimination of hedonic contrast for the experts and their rating the test orchids equivalently to Group C.

General Discussion

These studies confirm Fechner's principle of hedonic contrast (Beebe-Center, 1932/1965) and previous findings that categorization attenuates such contrast (Dolese et al. 2005; Zellner et al. 2002, Zellner et al. 2003). In fact, among the experts in Experiment 2 hedonic contrast was eliminated because they categorized the context iris and test orchids into different categories. Because the subordinate level is the basic level for experts, they perceive the two kinds of flowers to be in different categories and don't compare them. On the other hand, for the novices in Experiment 1 the two kinds of flowers are all members of their basic category of flowers and are thus compared.

The lack of a contrast effect in the present Experiment 2, when experts were judging the attractiveness of test orchids following context iris (Group IO), is the opposite of what is predicted by Schwarz and Bless (1992). They predict that because experts divide the objects with which they are expert into more categories (i.e., they use what is normally the subordinate level as their basic level) more rather than fewer contrast effects should occur in these experts than in novices (p.232). However, the reason they predict such an outcome is because the contrast effects they discuss are entirely different from the ones we are examining which result from comparison processes. The contrast effects they study are not the result of subjects comparing stimuli within a category. Instead, they study the effect of the inclusion or exclusion of a category member in the evaluation of that category. In particular, they have studied the evaluations of political parties which either include or exclude a particular highly regarded member of a political party (e.g., Bless & Schwarz, 1998). If this highly regarded member is included in the party, the subjects judge the party more favorably (what the authors call assimilation) than when he is excluded (what they call contrast).

Any situation in which subjects are comparing members of the same category will result in contrast; contrast will be reduced when the stimuli being compared are in different categories. This kind of effect has been found with physical judgments (Brown, 1953; Coren & Enns, 1993; Stapel & Koomen, 1997) and social judgments (Aarts & Dijksterhuis, 2002; Stapel & Winkelman, 1998) as well as hedonic judgments. In judgments of these types resulting from comparison processes, experts will experience less contrast within their area of expertise than will novices. Experts have developed an extensive system of subcategorizations with which they deal with items in their area of expertise. They thus make fewer intra-category comparisons than do people who place all stimuli into only a few categories.

References

Aarts, H. & Dijksterhuis, A. (2002). Category activation effects in judgment and behaviour: The moderating role of perceived comparability. *British Journal of Social Psychology, 41,* 123-139.

Beebe-Center, J.G. (1965). *The psychology of pleasantness and unpleasantness.* New York: Russell & Russell. (original work published 1932).

Bless, H. & Schwarz, N. (1998). Context effects in political judgement: assimilation and contrast as a function of categorization processes. *European Journal of Social Psychology, 28,* 159-172.

Brown, D.R. (1953). Stimulus-similarity and the anchoring of subjective scales. *American Journal of Psychology, 66*, 199-214.

Coren, S. & Enns, J.T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics, 54*, 579-588.

Dolese, M.J., Zellner, D.A., Vasserman, M. and Parker, S. (2005). Categorization affects hedonic contrast in the visual arts. *Bulletin of Psychology and the Arts, 5,* 21-25.

Fechner, G.T. (1898). *Vorschule der Aesthetik II* (2nd ed.). Leipzig: Breitkopf & Hartel.

Johnson, K.E. & Mervis, C.B. (1997). Effects of varying levels of expertise on the basic level of categorization. *Journal of Experimental Psychology: General, 126*, 248-277.

Hollander, M., & Wolfe, D. (1973). *Nonparametric Statistical Methods.* New York: John Wiley & Sons.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M. & Boyes-Braem. P. (1976). Basic objects in natural categories. *Cognitive Psychology, 7*, 382-439.

Schwarz, N. & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L.L. Martin and A. Tesser (Eds.) *The construction of social judgments*, pp. 217-245. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stapel, D.A. & Koomen, W. (1997). Social categorization and perceptual judgment of size: When perception is social. *Journal of Personality and Social Psychology, 73*, 1177-1190.

Stapel, D.A. & Windielman, P. (1998). Assimilation and contrast as a function of contest-target similarity, distinctness, and dimensional relevance. *Personality & Social Psychology Bulletin, 24*, 634-646.

Tanaka, J.W. & Taylor, M. (1991). Object categories and expertise: is the basic level in the eye of the beholder? *Cognitive Psychology, 23*, 457- 482.

Zellner, D.A., Kern, B.B., & Parker, S. (2002). Protection for the good: subcategorization reduces hedonic contrast. *Appetite, 38*, 175-180.

Zellner, D.A., Rohm, E.A., Bassetti, T.L. & Parker, S. (2003). Compared to what? Effects of categorization on hedonic contrast. *Psychonomic Bulletin & Review, 10*, 468-473.

EFFECT OF SPACE ON AUDITORY TEMPORAL PROCESSING WITH A SINGLE-STIMULUS METHOD

Martin Roy and Simon Grondin
Université Laval, Québec, Canada

ABSTRACT

In the present experiment, participants had to categorize, as short or long, brief temporal intervals presented according to a single-stimulus method. The intervals were marked by two brief auditory signals delivered from two of four different speakers. There was a distance of 1.1 or 3.3 m between each pair of speakers. The main finding is that, for both ranges of durations under investigation (125 and 250 ms), the greater the space between signals, the shorter the perceived duration, which is inconsistent with the so-called kappa effect. Moreover, increasing the distance between the marker sources also resulted in an increased Weber fraction. These findings can be accounted for by the role of attention in an internal clock consisting of a "pacemaker-counter" device.

In order to study timekeeping mechanisms, researchers often ask participants to judge the relative length of temporal intervals (Grondin, 2001). These intervals are most often marked by two brief sensory events (auditory, visual, tactile) delivered from a unique source. The question addressed here is related to a special case where, instead of using a unique source to deliver signals marking an interval, signals were delivered from different spatial locations. Auditory signals were used in the present study.

In psychology, a so-called *kappa effect* is known to exist in space-time relations: duration is perceived as longer when there is more space between signals marking time (Jones & Huang, 1982). This effect has been tested more often in the visual mode (Cohen, Hansel & Sylvester, 1953, 1955; Collyer, 1977; Miyatani, 1984-1985), and even in the tactile mode (Suto, 1952, 1955, 1957). There is only limited evidence for this effect in the auditory mode (Cohen, Hansel, & Sylvester, 1954; Jones & Huang, 1982; Yoblick & Salvendy, 1970; see Ouellet, 2003).

The *kappa effect* is usually shown to be robust in conditions where three successive signals (say, X, Y, and Z, with Y somewhere in between X and Z) are delivered. For two equal time intervals defined by the onset of two signals, X-Y or Y-Z, duration is perceived as longer for the X-Y sequence than for the Y-Z one if the distance between X and Y is greater than the distance between Y and Z. On the other hand, it has been shown that, in the visual mode, there are cases where duration can be perceived as shorter when space between signals is increased (Guay & Grondin, 2001). This result was observed in an experiment where three signals (**A**bove, **M**iddle and **B**elow) were located in front of participants on the same vertical plane. Temporal intervals were more often perceived as longer when they were marked by a B-A or A-B sequence, in comparison with intervals marked by an M-B, B-M, M-A or A-M sequence. Another important aspect of this experiment is that a single-stimulus method (one categorization judgment after the presentation of one interval) and marker-type intervals were randomized within blocks of trials.

The purpose of the present study was to verify if space exerts influence on temporal estimation (1) when intervals to be estimated are marked by sounds delivered from sources having different distances between them, and (2) when these intervals are presented according to a single-stimulus method, with randomized presentations of the marker-type conditions.

Method

Participants

Twelve 19- to 26-year-old volunteer students at Université Laval (six females and six males) with no hearing problems participated in this experiment. They were paid CAN $20 ($2.50 per session) for their participation.

Apparatus and stimuli

The intervals to be discriminated were a silent duration between two 20-ms auditory stimuli. The stimuli were a 1-kHz pure sinusoidal sound generated by IBM PC running E-Prime software (version 1.1.4.1 - SP3). The computer was equipped with an SB Audigy 2 sound card, and the sounds were delivered by Logitech Z-640 loudspeakers. Participants pressed ‘‘1’’ or ‘‘3’’ on the computer keyboard to indicate that the interval was short or long, respectively.

Procedure

The single-stimulus method was employed (Allan, 1979; Morgan, Watamaniuk & McKee, 2000) i.e., each trial consisted in presenting one interval. Participants were asked to judge whether the time interval between the two sensory signals belonged to the ‘‘short’’ or to the ‘‘long’’ category of a given distribution of intervals around a mid-point value (base duration). A 1.5-s feedback signal was presented immediately after the response on the computer screen and indicated whether the response was correct or not.

There were two base durations, 125 and 250 ms. In the former case, short intervals lasted 104, 110, 116 and 122 ms, and long intervals 128, 134, 140 and 146 ms. In the latter case, short intervals lasted 208, 220, 232 and 244 ms, and long intervals 256, 268, 280 and 292 ms. There were two distances between the auditory sources (speakers), 1.1 m and 3.3 m (see Figure 1).

Twelve participants completed eight sessions, four at 125 ms and four at 250 ms. Six participants completed the sessions at 125 ms before 250 ms, and six completed their sessions in the reverse order. There were six blocks of 64 trials in each session. There were eight repetitions, in a random order, of each of the eight intervals (four short and four long intervals) within each block.

Data analysis

For each participant and for each of the four conditions (2 distances and 2 base durations), an 8-point psychometric function was traced, plotting the eight empty intervals on the *x-axis* and the probability of responding “long” on the y-*axis*. Each point on the psychometric function was based on 96 presentations.

The *pseudo-logistic model* (Killeen, Fetterman & Bizo, 1997) was fitted to the resulting curves. Two indices of performance were estimated from each psychometric function, one for sensitivity and one for perceived duration. As an indicator of temporal sensitivity, estimates of one standard deviation (SD) on the psychometric function were determined. Using one SD (or

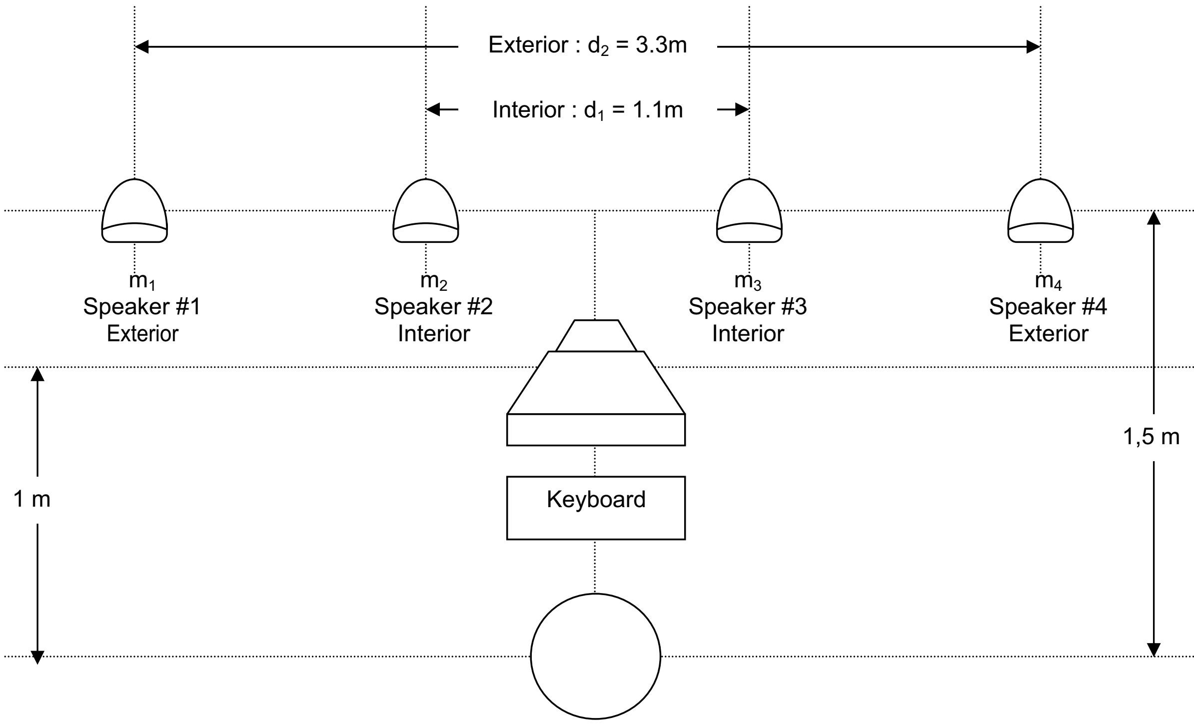


Figure 1. Experimental set-up

variance) is a common procedure to express temporal sensitivity (Grondin & Rammsayer, 2003; Grondin, Roussel, Gamache, Roy & Ouellet, 2005; Killeen & Weiss, 1987).

The other parameter estimated was the temporal bisection point (BP). The BP can be defined as the *x* value corresponding to the 0.50 probability of "long" responses on the *y*-axis. The observed shift of the BP for different conditions can be interpreted as an indication of differences in perceived duration. Thus, longer perceived durations are reflected by smaller BP values.

Results

In order to allow direct comparisons between base durations, two dependent variables were derived from the estimated parameters. One is the *Constant Error*, which is the BP minus the base duration. The other is the *Coefficient of Variation*, which is the SD divided by the BP.

Figure 2 shows the results for the *Constant Error*. Essentially, it reveals higher values in the 3.3-m condition than in the 1.1-m condition. A 2 x 2 ANOVA with repeated measures revealed that both the distance, $F(1,11) = 10.54$, $p < .01$, and the standard duration effects, $F(1,11)= 4.91$, $p < .05$, were significant. The interaction effect was also significant, $F(1,11) = 8.35$, $p < .05$.

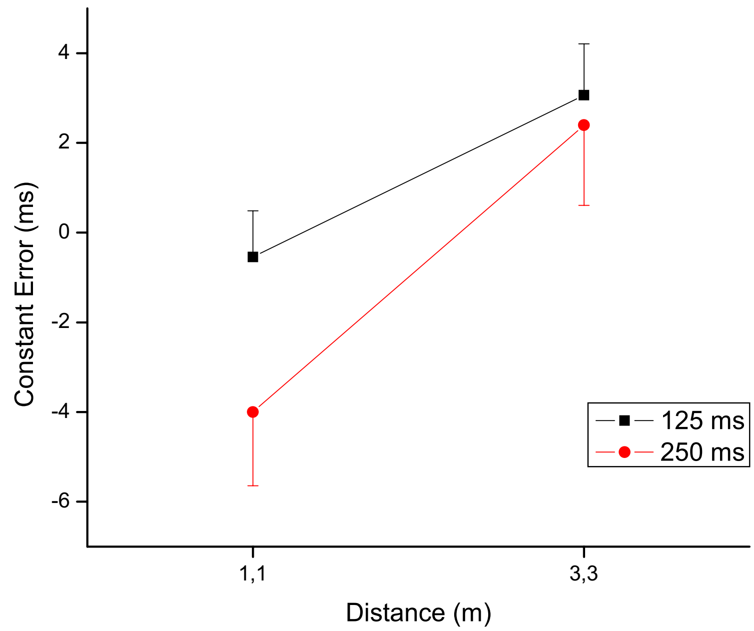


Figure 2. **Mean *Constant Error* as a function of distance between auditory markers. (Bars are standard errors)**

Figure 3 shows the results for the *Coefficient of Variation*. The 2 x 2 ANOVA with repeated measures revealed that both the distance, $F(1,11) = 9.71$, $p < .01$, and the standard duration effect, $F(1,11) = 19.28$, $p < .01$, were significant. The interaction effect was not significant, $F(1,11) = .85$, $p = .38$.

Discussion

The results of the present experiment clearly indicate that increasing the distance between sound sources marking time intervals leads to a decrease of the perceived duration (a higher constant error). These results are inconsistent with what is usually reported when referring to the *kappa effect* but consistent with results obtained in the visual mode with a single-stimulus method (Guay & Grondin, 2001). Other results linking time and space in the auditory mode revealed no such effect of distance between sound sources when sequences of four sounds from four sources were used (Ouellet, 2003). The present experiment also revealed that increasing distance between marker's sources results in a higher coefficient of variation.

The present results can perhaps be explained on the basis of a classical model of timing. Most contemporary researchers in the field of time perception refer to an internal-clock hypothesis. Such a central clock is usually described as a pacemaker-counter device, with the first structure emitting pulses accumulated by the second one (Grondin, 2001; Killeen & Weiss, 1987). It is this accumulation that forms the basis on which time is estimated. This accumulation is also under the control of an attentional mechanism, with more attention to time resulting in a

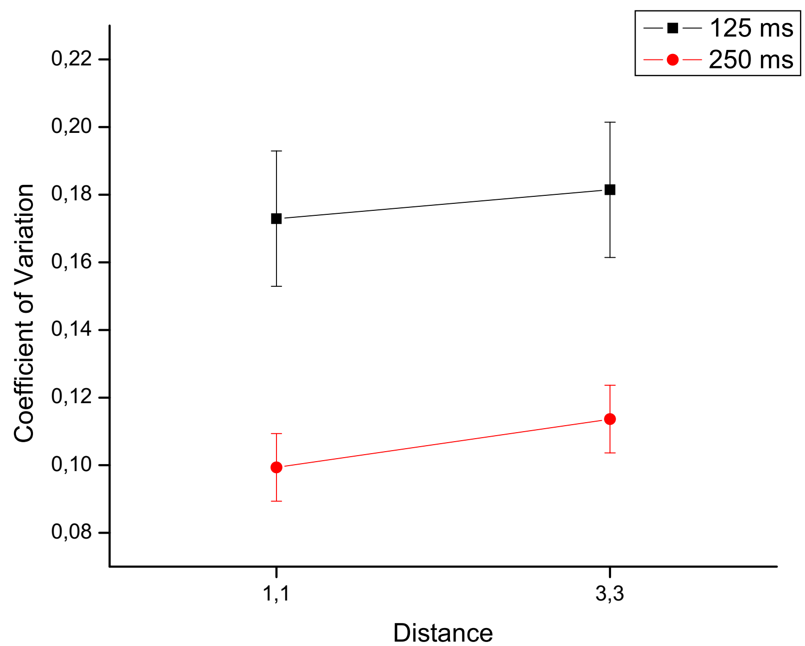


Figure 3. Mean *Coefficient of Variation* as a function of distance between auditory markers. (Bars are standard errors)

higher accumulation of pulses (Grondin & Macar, 1992; Macar et al., 1994).

That more space between sources leads to shorter perceived duration is consistent with the recognised role of attention in the clock-counter models of time, with a larger attentional displacement for one source to another resulting in less attention to time and therefore in a greater loss of pulses. Moreover, this explanation is also consistent with the results obtained with the coefficient of variation. Disturbing attention with space causes more variance (more categorization errors – higher coefficient of variation) in the accumulation process.

Finally, the results also revealed higher coefficients of variation at 125 than 250 ms. This finding is consistent with a generalized form of Weber's law applied to time in which sensory noise (nontemporal noise due to attention disturbance) causes more damage to performance with briefer intervals.

References

Allan, L. G. (1979). The perception of time. *Perception & Psychophysics, 26*, 340-354.

Cohen, J., Hansel, C.E.M., & Sylvester, J.D. (1953). A new phenomenon in time judgment. *Nature, 172*, 901.

Cohen, J., Hansel, C.E.M., & Sylvester, J.D. (1954). Interdependence of temporal and auditory judgments. *Nature, 174*, 642-644.

Cohen, J., Hansel, C.E.M., & Sylvester, J.D. (1955). Interdependence in judgments of space, time and movement. *Acta Psychologica, 11*, 360-372.

Collyer, C.E. (1977). Discrimination of spatial and temporal intervals defined by three light flashes : Effects of spacing on temporal judgments and of timing on spatial judgments. *Perception & Psychophysics, 21,* 357-364.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127,* 22-44.

Grondin, S., & Macar, F. (1992). Dividing attention between temporal and nontemporal tasks: A performance operating characteristic -POC- analysis. In F. Macar, V. Pouthas & W. Friedman (Eds.), *Time, Action, Cognition: Towards Bridging the Gap*. (pp. 119-128). Dordrecht, Netherlands: Kluwer.

Grondin, S., & Rammsayer, T. (2003). Variable foreperiods and duration discrimination. *Quarterly Journal of Experimental Psychology, 56A*, 731-765.

Grondin, S., Roussel, M.-È., Gamache, P.-L., Roy, M., & Ouellet, B. (2005). The structure of sensory events and the accuracy of judgments about time. *Perception*, *34*, 45-58.

Guay, I., & Grondin (2001). Influence on time interval categorization of distance between markers located on a vertical plane. In E. Sommerfeld, R. Kompass, & T. Lachman (Eds.), Proceedings *of the 17th Annual Meeting of the International Society for Psychophysics* (pp. 391-396). Berlin: Pabst Science Publishers.

Jones, B., & Huang, Y.L. (1982). Space-time dependencies in psychophysical judgment of extent and duration : Algebraic models of the tau and kappa effect. *Psychological Bulletin, 91,* 128-142.

Killeen, P.R., Fetterman, J.G., & Bizo, L.A. (1997). Time's cause. In C.M. Bradshaw & E. Szabadi (Eds). Time and Behavior : Psychological and Neurobehavioral Analyses (pp. 79-131). Amsterdam, Netherlands : North-Holland/Elsevier Science.

Killeen, P. R., & Weiss, N. A. (1987). Optimal timing and the Weber function. *Psychological Review, 94*, 455-468.

Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition, 22*, 673-686.

Morgan, M.J., Watamaniuk, S. N. J., McKee, S. P., (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, *40*, 2341-2349.

Miyatani (1984-1985). The time and distance judgments at different levels of discriminability of temporal and spatial information. *Hiroshima Forum for Psychology, 10,* 45-55.

Ouellet, B. (2003). *L'influence de la distance entre des marqueurs statiques sur la discrimination d'intervalles temporels : À la recherche de l'effet kappa classique en modalité auditive.* Unpublished master's dissertation, Université Laval, Québec, Canada.

Suto, Y. (1952). The effect of space on time estimation (S-effect) in tactual space. I. *Japanese Journal of Psychology, 22,* 45-57.

Suto, Y. (1955). The effect of space on time estimation (S-effect) in tactual space. II : The role of vision in the S effect upon skin. *Japanese Journal of Psychology, 26,* 94-99.

Suto, Y. (1957). Role of apparent distance in time perception. *Research Reports of Tokyo Electrical Engineering College, 5,* 73-82.

Yoblick, D. A., & Salvendy, G. (1970). Influence of frequency on the estimation of time for auditory, visual, and tactile modalities: The kappa effect. *Journal of Experimental Psychology, 86,* 157-164.

Acknowledgements. This research was made possible by a grant awarded to SG by the Natural Sciences and Engineering Council of Canada, and NSERC. We would like to thank Marie-Claude Simard and Karine Drouin for their help with data collection. Correspondence should be addressed to Simon Grondin, École de psychologie, Université Laval, Québec, Qc, Canada, G1K 7P4 (E-mail: simon.grondin@psy.ulaval.ca)

ATTENTIONAL ALLOCATION AND SUBJECTIVE CONFIDENCE CALIBRATION

Jordan Schoenherr, Dr. Craig Leth-Steensen, and Dr. William M. Petrusic
bold_resolution@hotmail.com, craig_leth_steensen@carleton.ca, bpetrusi@carleton.ca
Carleton University, Ottawa

Abstract

Twenty participants made perceptual judgements involving bi-modal, bi-dimensional stimuli under various attentional allocation conditions, both with and without providing confidence judgements. Decreasing attentional allocation to a stimulus modality affected both RTs, accuracy, and confidence calibration measures. Rendering confidence increased both RT and accuracy.

The nature of the calibration of confidence in perceptual judgment is only beginning to be fully explored in a contemporary fashion (Baranski & Petrusic, 1994). In a seminal paper, Baranski and Petrusic (1994) collected ratings of percent confidence for perceptual judgments (e.g., which of two small vertical lines was either nearer or farther to a central midline) and compared them to the actual, observed accuracies of the associated responses. Contrary to currently accepted opinion at the time (c.f., Bjorkman, Juslin, & Winman, 1993), they found that overconfidence can be obtained for perceptual judgments when the difficulty of those judgements is enhanced either explicitly by using very confusable perceptual stimuli or implicitly by using conditions emphasizing very speeded responding. Such a finding mirrored the well-known "hard-easy" calibration difficulty effect obtained in research involving subjective probability assessments of intellectual knowledge (Gigerenzer, Hoffrage, & Kleinbolting, 1991) and provided a basis of support for the view that the mechanisms underlying the calibration of confidence are similar for both perceptual and non-perceptual judgments. Baranski and Petrusic (1994) also found that both calibration and resolution confidence indices can worsen when judgment difficulty is enhanced explicitly. Moreover, although the provision of accuracy feedback did not significantly affect calibration in their study, it did improve resolution (especially for the easier judgments, Baranski & Petrusic, 1994; although see also Petrusic & Baranski, 1997, where such feedback was shown to both increase overconfidence and either degrade or improve calibration depending on the easiness or hardness, respectively, of the surrounding contextual difficulty).

In our present study, two perceptually based stimuli were presented consecutively on each decision trial. These stimuli themselves were compound stimuli made up of simultaneously presented visual (squares that differed in terms of both size and shading) and auditory (pure tones that differed in terms of both pitch and loudness) stimuli. In the spirit of a classic study on multidimensional discrimination ability by Lindsay, Taylor, and Forbes (1968), after presentation of the second compound stimulus, participants were cued to make a judgment regarding the nature of the change that occurred on one of the four possible stimulus attributes. Decision difficulty was enhanced by manipulating the amount of attention to-be-allocated to each (visual and auditory) perceptual modality. As is well-known, attention can be divided along multiple perceptual channels without the total cessation of the processing of stimuli in "unattended" channels (Treisman, 1960); however, such information is processed less efficiently. Our study extends that previous work by allowing for an explicit examination of the role of attention on confidence processing. Given the metacognitive nature of confidence processing, it seems likely that it should be susceptible to moderation by the amount of attention allocated to stimulus processing.

In addition, both decision response times (DRTs) and confidence judgment response times (CRTs) were obtained in our study allowing for an examination of the relation between both kinds of RTs and the levels of reported confidence. In general, a decrease in DRT with increasing confidence level is a finding that has been well-documented (Baranski & Petrusic, 1994; although, as demonstrated by them, such RTs can also depend on explicit decisional difficulty at each confidence level which then rules out the possibility that confidence is derived from a direct scaling of DRT). In contrast, the study of the time to provide the confidence judgements themselves is a fairly recent enterprise initiated by Baranski and Petrusic (1998). In Baranski and Petrusic (1998), it was found that CRTs, although much shorter overall than DRTs, varied across confidence levels but not explicit decisional difficulty levels (but more so under speed emphasis and at lower levels of practice; although see also Baranski & Petrusic, 2001, where analogous CRT effects occurred for very well-practiced subjects under a pure accuracy-emphasized response set). This finding strongly implies that there is a post-decisional, computational component to confidence processing.

Finally, in our study, participants performed blocks of perceptual judgment trials where confidence judgements were required and identical blocks of trials where confidence was not required. This manipulation was examined because Petrusic and Baranski (2003) have shown that the comparative decision process itself can be affected by the requirement to provide subsequent confidence judgements. In their study, two groups of participants compared the sizes of simultaneously presented squares. For one of the groups, those judgments were followed by confidence ratings. Petrusic and Baranski (2003) found that DRTs were much higher for the participants who provided confidence judgments than for those who did not (although, somewhat counterintuitively, this effect was larger for the explicitly easier pairs). In addition, using a within-participant manipulation of confidence versus no confidence rendering, Baranski and Petrusic (2001) obtained a similar inflation of DRT with confidence (which, however, did not depend on the level of explicit decisional difficulty). In neither of these two studies (Baranski & Petrusic, 2001; and Petrusic & Baranski, 2003) was decision accuracy significantly affected by the requirement to provide confidence.

Method

Twenty participants performed perceptual judgments for 90 min. The bi-modal and bi-dimensional stimuli consisted of squares varying in size and shade presented simultaneously with pure tones varying in pitch and loudness. Participants compared each square/tone compound stimulus with a second successively presented one (each presented for 500 ms with a 500 ms inter-stimulus interval), indicating manually after an immediate response cue whether the first or second stimulus was either the larger, the darker, the higher, or the louder. A complete listing of the stimulus pairs used is given in Table 1 (note that for each modality each type of stimulus pair was presented equally often in each block, randomly matched with pairs from the other modality, and also that there was an attempt, through pilot study work, to adjust the level of differences in order to neutralize any potential time-order errors). Accuracy feedback was not provided.

There were five attentional manipulation conditions of 64 trials each within which the participants were instructed to allocate either all of their attention to only one of the modalities (100-0% and 0-100%), a majority of their attention to one modality and a minority to the other (75-25% and 25-75%), or half of their attention to each modality (50-50%). These were also the percentage of times that the corresponding modality was then cued for response in that block (each associated stimulus dimension was then cued equally often). After two 100% allocation training blocks, participants began with either a 100% auditory or visual modality allocation block (whose order was counterbalanced across participants), gradually reducing the attentional allocation to that modality (and increasing it to the other modality) as the blocks progressed.

Each participant performed all five attentional condition blocks both with and without confidence judgements (half providing confidence in the first five blocks and half providing it in the second five blocks). Percent confidence was indicated by way of six keys labeled 50, 60, 70, 80, 90, and 100, respectively. Confidence performance measures were: over/underconfidence (the difference between the mean confidence rating and the mean proportion correct), calibration (the degree of correspondence between confidence ratings and actual proportion correct), resolution (the degree to which proportion correct differs across confidence rating categories), and normalized resolution (□□^{2}).

Table 1: Stimuli

Presentation Order			
Auditory		**Visual**	
First	***Second***	***First***	***Second***
1000_65	980_58	**60_120**	48_108
	980_70		48_140
	1020_58		67_108
	1020_70		67_140
980_58	**1000_65**	53_92	**60_120**
980_70		53_132	
1020_58		72_92	
1020_70		72_132	
500_55	490_49	**30_160**	27_142
	490_60		27_184
	510_49		35_142
	510_60		35_184
490_49	**500_55**	25_134	**30_160**
490_60		25_180	
510_49		33_134	
510_60		33_180	

Results

Repeated measures ANOVAs were performed using mean accuracy and correct DRT as the dependent variables (Tables 2 and 3 respectively; note that providing confidence first or second was also included as a between-subject nuisance variable in all of the following analyses but did not enter into any significant main effects or interactions). For accuracy, significant main effects were observed for confidence rendering ($F(1, 18) = 7.118$, $p < .016$), stimulus modality ($F(1, 18) = 32.316$, $p < .00002$), and attention condition ($F(3, 54) = 11.224$, $p < .00001$). The Stimulus Modality x Attention Condition interaction was also significant ($F(3, 54) = 2.895$, $p < .043$). For DRT, significant main effects of confidence rendering ($F(1, 18) = 6.363$, $p < .021$), stimulus modality ($F(1, 18) = 18.292$, $p < .00045$), and the attention condition linear trend ($F(1, 18) = 5.181$, $p < .035$) only were found.

Table 2 : Averaged Accuracy by Attention Condition

		25	50	75	100
NC	AUD	.6915	.6550	.7223	.7376
	VIS	.7799	.8325	.8446	.8638
	M	**.7357**	**.7441**	**.7837**	**.8007**
CON	AUD	.7010	.7062	.6989	.7707
	VIS	.8194	.8745	.8792	.8811
	M	**.7602**	**.7903**	**.7890**	**.8259**

Table 3: Averaged $DRT_{Correct}$ by Attention Condition

		25	50	75	100
NC	AUD	1899	1824	1782	1450
	VIS	1362	1258	1123	1037
	M	**1630**	**1541**	**1452**	**1244**
CON	AUD	2339	2130	3315	1829
	VIS	1647	1478	1441	1387
	M	**1993**	**1804**	**2378**	**1608**

Table 4: Averaged Calibration Indices for the Confidence Blocks

Level	p(correct)	M_{Conf}	O/U	Cal	Res	[illegible]^{2}
25	.7602	.7670	.0068	.0353	.0433	.2541
50	.7903	.7973	.0070	.0192	.0349	.2202
75	.7890	.8056	.0166	.0189	.0284	.1779
100	.8259	.8369	.0110	.0135	.0910	.6519
M	.7919	.8023	.0104	.0217	.0494	.3263

Repeated measures ANOVAs involving the calibration indices (Table 4 and see Figure 1 for the group calibration curves) revealed significant effects for calibration ($F(3, 54) = 9.922$, $p < .0003$), resolution ($F(3, 54) = 21.182$, $p < .00001$), and normalized resolution ($F(3, 54) = 25.756$, $p < .00001$) across attention condition. Mean confidence also varied across attention condition ($F(3, 54) = 13.442$, $p < .0001$) and stimulus modality ($F(1, 18) = 4.34$, $p < .052$)

In addition, both DRT ($F(5, 90) = 19.895$, $p < .00001$) and CRT ($F(5, 90) = 10.061$, $p < .00001$) differed significantly across confidence rating categories (see Figure 2).

Discussion

In general, performance deteriorated significantly with division of attention for all of the dependent variables except over/underconfidence. One caveat is that because accuracy also decreased with decreasing attentional allocation, it is not clear whether the worsening of both calibration and resolution represents an attentional effect or simply a general difficulty effect (c.f., Baranski & Petrusic, 1994). Another important finding was that, as in both Baranski and Petrusic (2001) and Petrusic and Baranski (2003), DRT, and importantly decision accuracy here as well, was significantly higher for the blocks in which confidence judgments were being rendered. This finding suggests either that (a) the requirement to provide confidence slows but also sharpens the decision process, or that (b) responding simply occurs more cautiously when subsequent confidence judgments are required (however, note that this still represents an adjustment of the decision process). Finally, as a further replication of Baranski and Petrusic (1998), both DRT and CRT significantly varied with confidence rating category.

Figure 1: Aggregate Confidence Calibration

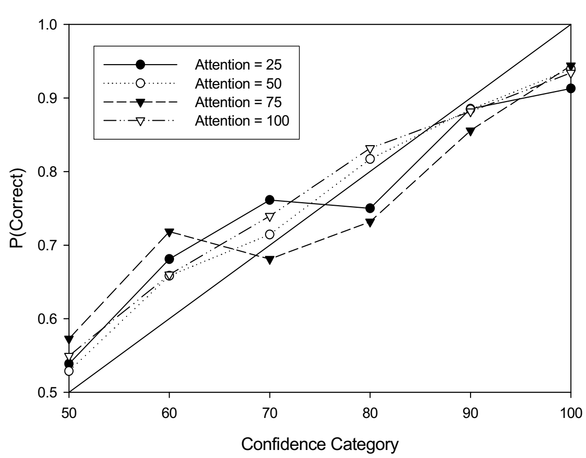
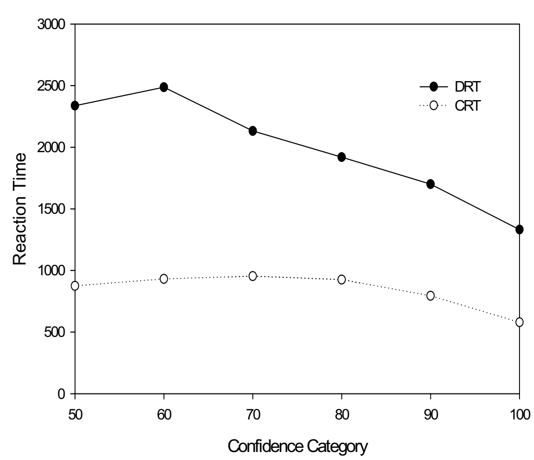


Figure 2: Decision and Confidence RTs



Note that for the group calibration curves, over/under confidence was .0052, .0052, .0141, and .0111, calibration was .0041, .0019, .0044, and .0027, and resolution was .0198, .0216, .0171, and .0627 for the 25, 50, 75, and 100% attention conditions respectively.

References

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception & Psychophysics*, *55*, 412-428.

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgements: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 929-945.

Baranski, J. V., & Petrusic, W. M. (2001). Testing architectures of the decision-confidence relation. *Canadian Journal of Experimental Psychology*, *55*, 195-206.

Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics*, *54*, 75-81.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabalistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506-528.

Linday, P. H. Taylor, M. M., & Forbes, S. M. (1968). Attention and multidimensional discrimination. *Perception and Psychophysics*, *4*, 113-117 .

Petrusic, W. M., & Baranski, J. V. (1997). Context, feedback, and the calibration and resolution of confidence in perceptual judgements. *American Journal of Psychology*, *110*, 543-572.

Petrusic, W. M., & Baranski, J. V. (2003). Judging confidence influences decision processing in comparative judgements. *Psychonomic Bulletin & Review*, *10*, 177-183.

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology*, 12, 242-248.

TRACKING AND RESPONDING TO A VISUAL TARGET

Yasuhiro Seya and Shuji Mori
Department of Kinesiology, Graduate School of Science, Tokyo Metropolitan University, 1-1 Minami-Ohsawa, Hachioji, Tokyo 192-0397, Japan

Abstract

To investigate effects of retinal eccentricity on reaction times (RTs) during smooth pursuit, we measured the RTs as a function of retinal eccentricity and of stimulus velocity. In Experiment 1, participants pursued a moving row of circular frames and responded to a target presented within one of the frames. In Experiment 2, the participants were presented with stimuli similar to those used in Experiment 1 while fixating on stationary central points. Experiment 1 showed that the RTs during smooth pursuit increased with increasing the retinal eccentricity. Although the RTs steeply increased from 5 to 10 deg/sec velocities, above 20 deg/sec velocities the RTs did not change systematically. Experiment 2 showed that the RTs during fixation increased with increasing the retinal eccentricity and the stimulus velocity. These results suggest that the RTs during smooth pursuit may have reflected a tradeoff between response speed (RT) and pursuit accuracy.

When an object moves through our visual field, we track it smoothly with our eyes. This eye movement is called smooth pursuit eye movement. Smooth pursuit plays an important role in maintaining a clear vision of the moving object. In everyday life, we are required to detect or discriminate the objects while tracking them with smooth pursuit.

Using reaction time (RT), many studies have examined temporal characteristics of visual information processing in detection and discrimination tasks. RTs are influenced by many factors, such as stimulus luminance (e.g., Bartlett & MacLeod, 1954) and retinal eccentricity (e.g., Rains, 1963). In a majority of those studies, participants were asked to keep their eyes on a stationary fixation point. However, relatively few studies have examined RTs during smooth pursuit. One such study is Van Donkelaar and Drew (2002). In their experiment, participants viewed a display containing a central cross ('×') and eight peripheral circles. The peripheral circles were placed horizontally at both sides of the cross, separated from each other by 1 deg. The cross and the circles moved rightward horizontally at 3, 5, 10 or 15 deg/sec, and after a randomized period one of the central cross or peripheral circles changed its shape (from '×' to 'o' or 'o' to '×'). The participants were instructed to pursue the cross accurately and to respond to the change of the shape as soon as possible. Van Donkelaar and Drew (2002) found that although RTs increased with increasing target eccentricity (a distance between the fixation stimulus and the target), the effect of the target eccentricity was attenuated as the stimulus velocity was increased. The results may imply that effect of retinal eccentricity (a distance between the target and the fovea) on RT decreased with increasing the velocity. Unfortunately, Van Donkelaar and Drew (2002) measured the RTs as a function of the target eccentricity, not the retinal eccentricity. Since eye velocity does not always match stimulus velocity during smooth pursuit, particularly at higher velocities (Murphy, 1978), the participants in Van Donkelaar and Drew's (2002) study may not have pursued the fixation stimulus accurately. It is thus possible that the target eccentricity used in their study may not have corresponded to the retinal eccentricity.

The main purpose of the present study was to investigate the effects of the retinal eccentricity on RTs during smooth pursuit. In Experiment 1, we measured the RTs during smooth pursuit as a function of the retinal eccentricity and stimulus velocity. In Experiment 2, we measured RTs during stationary fixation on a display of moving stimuli, which simulated the retinal image velocities of the stimuli presented in Experiment 1, in order to see whether the effect of retinal image velocity during smooth pursuit explained the RTs obtained in Experiment 1.

Experiment1

Method

Participants. Participants were five individuals (four males and one female with a mean age of 24.8 in the range of 23-26). One of them was the first author (male) and the others were naïve to the purpose of this study. All had normal or corrected-to-normal vision.

Apparatus and stimuli. A personal computer (Dell OptiPlex GX260) and color graphic system (Cambridge Research System VSG2/5) were used to control the experiment, RT measurement and recording from a response box. Stimuli were projected on a 17-inch high-resolution color display (TOTOKU CV722X) at a distance of 26 cm in front of the participants. Eye positions during the stimulus presentation were monitored with an infrared eye movement recording system (NAC EMR-8B). The temporal resolution of the system was 16.67 ms (sampling rate of 60 Hz) and the spatial resolution was 0.1 deg of visual angle.

In the stimulus display, seven circular-frames were presented in red (1.7 cd/m^2) on a black background (0.13 cd/m^2). Each frame subtended 0.8 deg in diameter. The frames were located horizontally, separated from each other by 5 deg, and the central frame served as the fixation stimulus. The frames were initially presented at 10 deg either to the right or the left sides of the center of the display, and moved horizontally back and forth at amplitude of 20 deg across the center of the display. The velocity of the frames was either 0, 1, 3, 5, 10, 15, 20, 30 or 40 deg/sec. Note that at 0 deg/sec velocity, the frames remained stationary at the center of the display. A circle, presented in white (7.5 cd/m^2), served as a target stimulus, subtending 0.37 deg in diameter. The target was presented within one of the frames (0, ± 5, ± 10, ± 15 deg from the center frame), with an equal probability of occurrence. The target moved at the same velocity and direction as the frames.

Procedure. The experiment was conducted in a dark booth. Participants were seated with their head fixed by a bite board and a chin rest while viewing the display monocularly (right eye). After dark adaptation for 5 minutes, participants were given several practice trials until they were familiarized with the task, followed by an experimental session. At the beginning of each trial, the frames were presented stationary. After 300 ms from the onset of the frames, the frames started to move. After a randomized foreperiod of 1-3 sec, the target was presented. The participants were asked to pursue the central frame accurately and to press a key as soon as the target appeared. The stimulus presentation was terminated by the key press and replaced with a blank display. The inter-trial interval was 2000 ms.

There were five sessions of 140 trials for each stimulus velocity. One session consisted of two blocks of 70 trials. The order of the velocity condition was randomized across the participants. All participants completed all the conditions over three or five days depending on their schedule and available time each day. They were given rest periods of 2 minutes between sessions.

Eye movement recordings and analysis. Horizontal eye positions during the stimulus presentation were recorded from the right eye of the participant. Calibration of the eye movement recordings was performed in the following way. First, before each stimulus velocity condition, the participants were asked to fixate nine points that were located at certain positions in the monitor of the recording system. These fixation data were used to calibrate the eye positions during the experiment. Second, in order to check the accuracy of the recordings, before and after the experimental session, the participants were also asked to fixate three points located horizontally at 0 and ± 10 deg from the center of the display. These fixation data were also used to calculate the retinal eccentricity of the target at its onset. All eye position data were stored on digital tape and analyzed off-line after the experiment. In order to analyze smooth pursuit, saccades that were detected by using a velocity criterion (above 90 deg/sec or 1.5 deg/frame) and eye movements during which the participant's eye moved in the opposite direction to the moving frames were removed from subsequent analysis. Horizontal eye velocity was calculated by dividing a distance at which the participants pursued the frames during the interval of 200-1000 ms (after the frames started to move) by the time.

Results

In all trials, the retinal eccentricity was calculated by subtracting the eye position from the target position at the onset of the target presentation. RT data were classified into seven retinal eccentricities: N2.4~T2.4 deg (N indicates that the target was presented in a nasal direction from the fovea and T indicates a temporal direction), N2.5~N7.4 deg, T2.5~T7.4 deg, N7.5~N12.4 deg, T7.5~T12.4 deg, above N12.5 deg and above T12.5 deg. The RT data were pooled across sessions and averaged for each condition for each participant.

RTs. Figure 1 presents the mean RTs of the five participants as a function of the retinal eccentricity. The figure showed that the RTs during smooth pursuit increased with increasing the retinal eccentricity. At all the eccentricities, the RTs increased steeply from 5 to 10 deg/sec velocities, whereas the RTs did not change much below 5 deg/sec or above 10 deg/sec velocities. As indicated by error bars, there were large individual differences at the velocities above 20 deg/sec.

Figure 2a presents the RTs separately for the five participants at the fovea (N 2.5~T2.5 deg), as a function of the stimulus velocity. For all the participants, the RTs during smooth pursuit increased steeply from 5 to 10 or 10 to 15 deg/sec velocities. At the velocities above 20 deg/sec, there were large individual differences: For S2 and S3 RTs increased with increasing the velocity; for S4 and S5 RTs did not change with increasing the velocity; for S1 RTs decreased with increasing the velocity.

Retinal image velocities. Except for 0 deg/sec velocity, the retinal image velocity was calculated by subtracting the eye velocity from the stimulus velocity. Figure 2b presents the retinal image velocities of the five participants as a function of the stimulus velocity. For all the participants, the retinal image velocities increased with increasing the stimulus velocity.

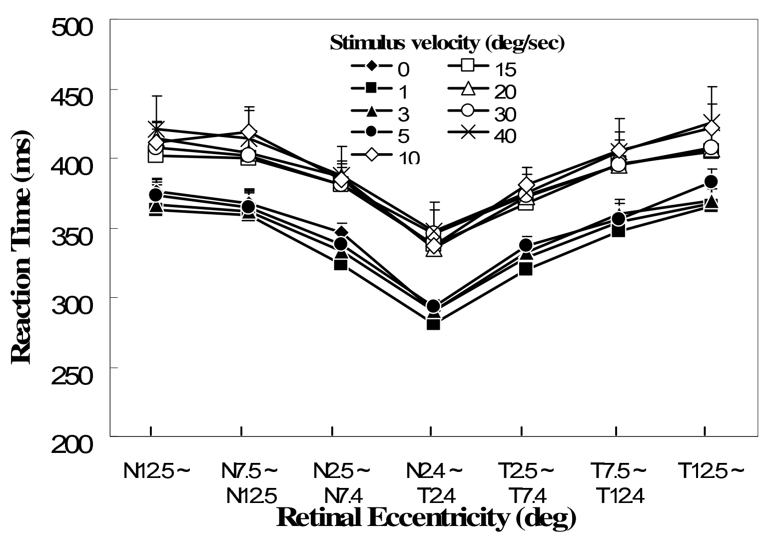


Figure 1. Mean RTs of the five participants as a function of the retinal eccentricity in Experiment 1. Vertical bars indicate standard errors of the mean. Note that N means that the target was presented in a nasal direction from the fovea, and T means a temporal direction.

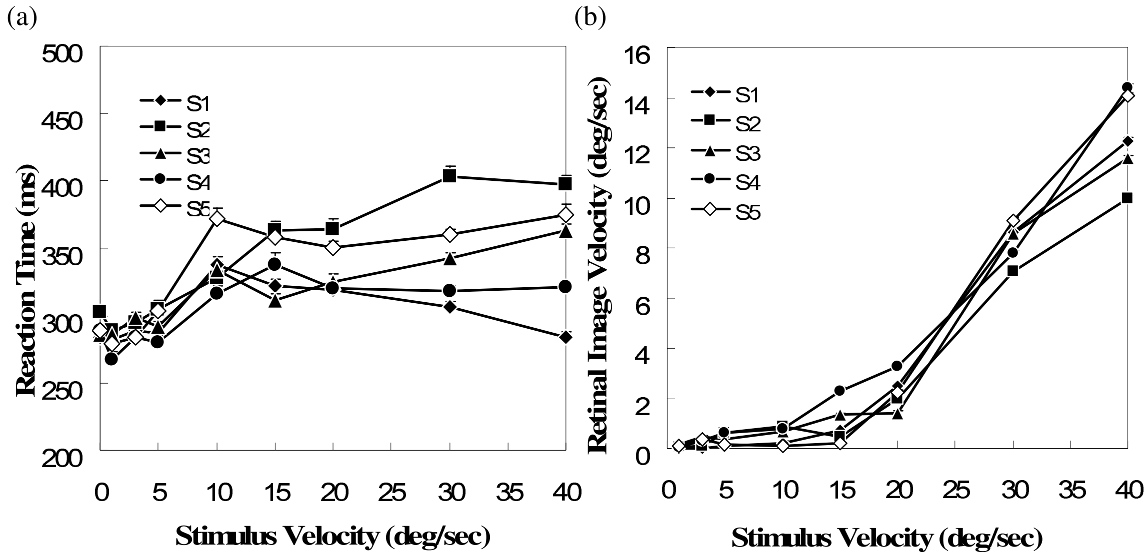


Figure 2. Individual data of RTs (a) and retinal image velocities (b) as a function of the stimulus velocity in Experiment 1. Vertical bars indicate standard errors.

Discussion

The results of this experiment showed that the RTs during smooth pursuit increased with increasing the retinal eccentricity. The effects of the retinal eccentricity did not change much with the stimulus velocity. These results suggest that, regardless of the stimulus velocity, the RTs during smooth pursuit, as well as those during fixation, increased with increasing the retinal eccentricity. Why were the effects of the target eccentricity used in Van Donkelaar and Drew's (2002) study, not the retinal eccentricity, attenuated with increasing the stimulus velocity? A possible reason is that, as we described in the Introduction, the participants in their study may not have pursued the fixation stimulus accurately, particularly at higher velocities. So, the target eccentricity may not have corresponded to the retinal eccentricity. For this reason, the effect of the target eccentricity may have seemed to be attenuated with increasing the stimulus velocity.

In this experiment, the RTs increased steeply from 5 to 10 deg/sec velocities, which was consistent with the result of Van Donkelaar and Drew's (2002) study. Above 20 deg/sec velocities, the RTs did not change systematically with increasing the stimulus velocity and there were large individual differences in the RTs. A possible reason for these changes with increasing the stimulus velocity is due to a tradeoff between response speed and pursuit accuracy. As shown in Figure 2, at 10 deg/sec velocity, all the participants pursued the frames with a very small retinal image velocity, almost identical to that at 5 deg/sec velocity, and the RTs increased steeply. Above 20 deg/sec velocities, for S1, S4 and S5, large retinal image velocities occurred, and the RTs decreased or did not change much with increasing the velocity. On the other hand, S2 and S3 pursued the frames with small retinal image velocities and the RTs increased. Therefore, some participants may have responded to the target quickly at the expense of the pursuit accuracy, while others may have preferred the pursuit accuracy over the response speed. As a result, large individual differences may have occurred.

Experiment 2

Method

There were two major changes from the methodologies used in Experiment 1. First, the moving frames and the target were presented while the participants fixated their eyes on a stationary triangular array of three points at the center of the display. The array subtended 1 × 1 deg, and a single point of the three points subtended 0.18 deg in diameter. Second, the velocity of the moving frames was either 0, 1, 3, 5, 10 or 15 deg/sec, in order to examine RTs during stationary fixation on a display of moving stimuli, which simulated the retinal image velocities of the stimuli presented in Experiment 1 (see Figure 2b). Participants were five individuals who had participated in Experiment 1. In all other aspects, the method was identical to that used in Experiment 1.

Results and Discussion

Figure 3 presents the mean RTs of the five participants as a function of the retinal eccentricity. The RTs during fixation increased with increasing the retinal eccentricity, like the result of Experiment 1. The RTs increased with increasing the stimulus velocity, which was partially consistent with the result of Experiment 1.

In order to clarify the differences between the RTs during smooth pursuit and those during fixation, Figure 4 presents the mean RTs in Experiment 1 and those in Experiment 2 at the fovea (a: T2.4~N2.4 deg) and the peripheral visual field (b: above T12.5 deg and above N12.5 deg) as a function of the retinal image velocity. At all the retinal image velocities, the RTs during smooth pursuit were larger than those during fixation, suggesting that the effect of the retinal image velocity during smooth pursuit did not explain the RTs in Experiment 1. A possible reason for these differences is due to an additional effort to pursue the moving frames accurately. When the participants pursued the moving frames, processing capacity (e.g., attention) may have been devoted to the pursuit of the moving frames, and as a result, the capacity devoted to responding to the target may have decreased, which may have caused larger RTs during smooth pursuit than those during fixation.

General Discussion

In this study, we investigated the effects of the retinal eccentricity on the RTs during smooth pursuit. In Experiment 1, the RTs during smooth pursuit increased with increasing the retinal eccentricity. Experiment 2 showed that the RTs during fixation increased with increasing the retinal eccentricity. These results suggest that, regardless of the stimulus velocity, the RTs during smooth pursuit, as well as those during fixation, reflect the effect of the retinal eccentricity of the target presentation.

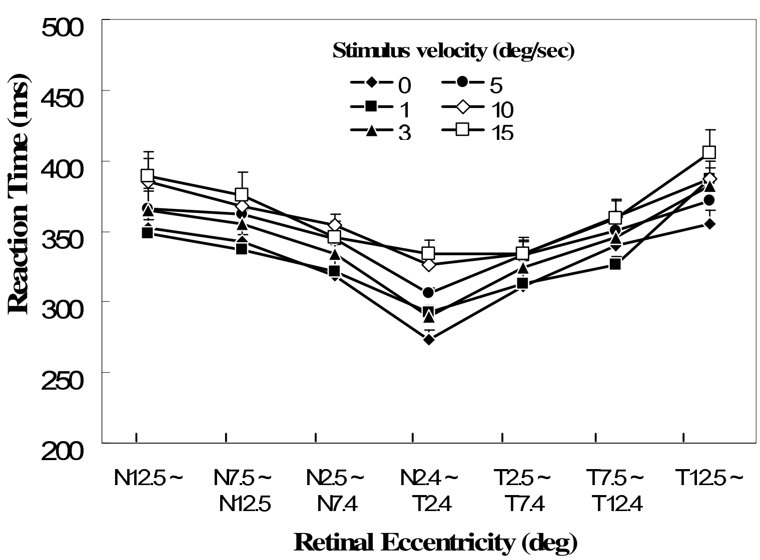


Figure 3. Mean RTs of the five participants as a function of the retinal eccentricity in Experiment 2. Vertical bars indicate standard errors. Note that N indicates that the target was presented in a nasal direction from the fovea and T indicates a temporal direction.

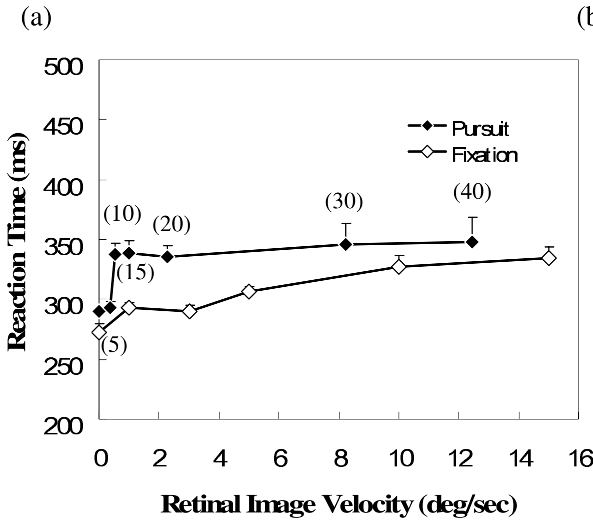


Figure 4. Comparison of the mean RTs of the five participants obtained in Experiments 1 and 2 for the fovea (a: N2.4~T2.4 deg) and peripheral visual field (b: above T12.5 deg and above N12.5 deg). Vertical bars indicate standard errors. Note that a value in parenthesis indicates the stimulus velocity (deg/sec) used in Experiment 1. N indicates that the target was presented in a nasal direction from the fovea and T indicates a temporal direction.

In Experiment 1, at all the retinal eccentricities the RTs during smooth pursuit increased steeply from 5 or 10 deg/sec velocities, but above 20 deg/sec velocities the RTs did not change systematically with increasing the velocity. These results may be explained by the tradeoff between the response speed and the pursuit accuracy, as we discussed in Experiment 1. However, in this study, we did not directly examine the tradeoff between the response speed and the pursuit accuracy. Further studies will be needed to explore this effect.

References

Bartlett, N. R., & MacLeod, S. (1954). Effect of flash and field luminance upon human reaction time. *Journal of the Optical Society of America, 44*, 306-311.

Murphy, B. J. (1978). Pattern thresholds for moving and stationary gratings during smooth eye movement. *Vision Research, 18*, 521-530.

Rains, J. D. (1963). Signal luminance and position effects in human reaction time. *Vision Research, 3*, 239-251.

Van Donkelaar, P. and Drew, A. S. (2002). The allocation of attention during smooth pursuit eye movements. In J. Hyönä, D. P. Munoz, W. Heide & R. Radach (Eds.), *Progress in Brain Research, 140*. Elsevier, Amsterdam, pp. 267-277.

EXTENDING THE SNARC EFFECT: SPATIAL COMPONENTS IN THE MENTAL REPRESENTATIONS OF NUMBERS, SYMBOLS, AND PERCEPTS

Samuel Shaki and William M. Petrusic
Department of Behavioral Sciences, College of Judea and Samaria, Ariel, 44837, Israel
Department of Psychology, Carleton University, Ottawa, K1S 5B6, Canada
samuel_shaki@hotmail.com; bill_petrusic@carleton.ca

Abstract

In each of three experiments requiring pairwise comparative judgments, a variety of SNARC effects were obtained. With comparison of the numerical magnitude of pairs of positive numbers, response times (RTs) were faster with the left hand than the right when the numbers were small but they were slower with the left hand than the right when the numbers were relatively large. This effect was obtained with both the instruction to select the smaller digit in the pair and the larger digit in the pair. Remembered symbolic comparisons with perceptual referents (e.g., size of animals) and with stimuli on well defined perceptual continua such as visual extent, in striking contrast to the SNARC effects with numbers, show a instruction dependent SNARC effect. With the instruction to select the stimulus lower on the underlying attribute, SNARC effects typically obtained with numbers occur. However, with the instruction to select the stimulus higher on the attribute, reverse SNARC effects were obtained.

It is well established that integers have both magnitude and spatial components in their mental representations. The most striking and influential demonstration of a *spatial* basis for the mental representation of a number line, with numbers arranged from left to right, was obtained by Dehaene, Bossini, and Giraux (1993). Dehaene et al. required their participants to classify the digits from zero to nine as odd or even; i.e., to make parity judgments. They found that parity judgments with the relatively small numbers in the set were faster with the left hand than with the right hand. However, when the numbers were relatively large, participants responded faster with their right hand. Capturing the association between the spatial component of the mental representation of numbers and the hand of the response, Dehaene et al. labeled their finding as the SNARC effect (Spatial Numerical Association of Response Codes).

Recently SNARC effects have also been obtained with well-defined highly over-learned linear orderings. For example, Gevers, Reynvoet, and Fias (2003) have shown, when participants are required to categorize a month of the year as coming before or after July, RTs are faster with the leftward than rightward responses with the earlier months in the year and conversely for the months later in the year. Similar effects were obtained in categorical judgements of whether a letter of the alphabet came before or after the letter M. Each of these experiments nicely shows that the mental representations of continua other than numbers have a spatial component. However, it remains to be determined if stimuli varying on other varieties of continua might also have spatial components in their mental representations.

Experiment 1

We begin this series of experiments by first establishing the properties of the SNARC effect with positive numbers using the method of pair-wise comparisons developed in Shaki and Petrusic (2005). Examining the factors controlling the occurrence of the SNARC effect with negative numbers, Shaki and Petrusic required comparative judgments of numerical magnitude using the instruction to choose the smaller number in the pair on a block of trials and the instruction to choose the larger on another block. Shaki and Petrusic found approximately comparable SNARC effects with the two instructions. The present experiment, thus, also offered an opportunity to determine if their findings would be replicated when the two instructions were also intermixed.

Method

Participants. Twelve Carleton University students participated in one 45 minute session to satisfy course requirements. All subjects reported normal or corrected-to normal vision.

Stimuli and Design. Four relatively small digits (0, 1, 2, 3) and four relatively large digits (6, 7, 8, 9) defined the two sets of digits used. The digits were paired within categories producing three relatively small digit pairs (0-1, 1-2, 2-3) and three relatively large digit pairs (6-7, 7-8, 8-9). Each pair in the design was presented in each of the two possible left-right position orders, resulting in 12 digit pairs. The two forms of the comparative instructions ("Smaller" "Larger") occurred equally often and were constant over a block in one condition (Blocked) and inter-mixed, occurring equally often and in random order in a second condition (Mixed). This factorial combination of the 12 stimulus pairs by two instructions by two conditions) was replicated ten times, preceded by three replications of practice trials.

Both the order in which the conditions were presented ("Blocked" first or "Mixed" first) and the order in which the two instructions were used (in the Blocked condition: smaller first then larger, or larger first then smaller) were counterbalanced. Precisely, the same sequence of blocks of trials was used for the practice trials as for the experimental trials for each participant. The order of presentation of the stimulus pairs within blocks was random and different for each participant.

Procedure. Participants were tested individually in a dimly lit room, seated approximately 80 cm from the centre of the video monitor. Participants were told that the presentation of either the word "SMALLER" or "LARGER" served as a warning for the next trial and was an instruction that indicated whether they were to choose either the larger or the smaller digit in the pair. After an additional 750 ms, the pair of digits appeared while the comparative instruction remained on the screen. The participant's task was to press the marked keyboard key ('A' on the left, and 'L' on the right) on the same side of the larger (or smaller, respectively) member of the pair. The presentation of the stimuli and the comparative instruction were response-terminated. The next trial began 1000 ms later. Participants were encouraged to respond quickly, but accurately.

The pairs of digits appeared at the respective centres of the left and right hemi-fields on the white background of a 17 inch (43 cm) View Sonic video monitor and the instructions, printed in David font (30, bold), appeared at the centre of the upper third of the screen. Event sequencing, randomization of trials and instructions, recording of responses and response times was under the control of SuperLab software run on a Pentium III microprocessor.

Results and Discussion

After combining the data over the two presentation orders, ANOVAs were conducted with mean correct RTs as the dependent variable and condition (blocked vs. intermixed), the six digit pairs, the two instructions, and hand of response as the within participant factors. In this experiment and the subsequent experiments, Huynh-Feldt, epsilon adjustment of degrees of freedom is used, although the degrees of freedom and error mean squares indicated in the text are those defined by the design. Level of significance is set at 0.05.

The ANOVA revealed that the stimulus pairs differed reliably ($F(5, 55)=16.39$, $MS(Error)=2777965.6$), with RTs generally increasing as the pairs increased in magnitude, reflecting the fact that numerical magnitude comparisons are based on relative magnitudes; i.e., a Weber Law like property. The pair by instruction interaction was also reliable ($F(5, 55)=5.98$, $MS(Error)=5126.7$) as a consequence of the typical, robust semantic congruity effects obtained with numerical magnitude comparisons.

More importantly, turning now to the effects involving the hand of the response; i.e., the SNARC effect with comparative judgements. As the plots in Figure 1 show, it is clear that RTs are faster with left hand responses than with right hand responses for the small number pairs, and faster with the right hand than with the left hand for the relatively large number pairs. The interaction involving stimulus pair and hand, which defines the SNARC effect with these paired comparisons is statistically reliable ($F(5, 55)=7.69$, $MS(Error)=2980.2$). Importantly, as is also evident in the plots in Figure 1, the obtained SNARC effects did not depend on instruction (F(5, 55)=2.08, MS(Error)=5477.7, p>.125) nor whether the instructions were blocked or randomized (F(5, 55)=.70, MS(Error)=2149.5, p>.529).

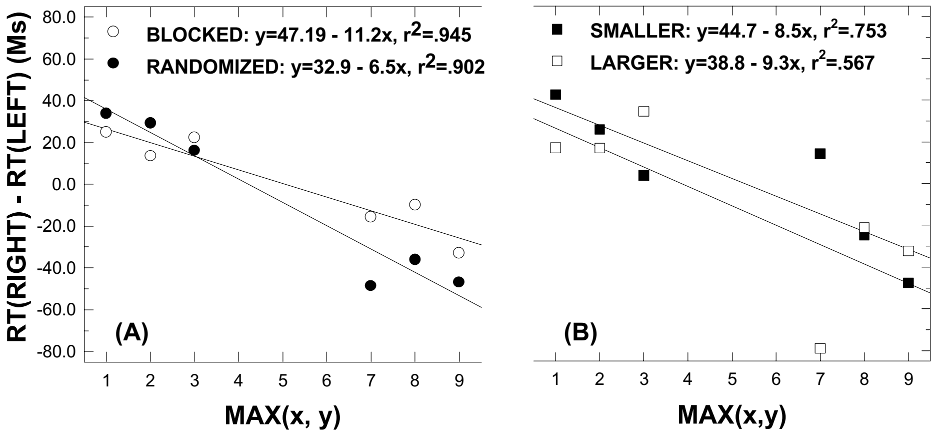


Figure 1. Mean RTs with the left hand subtracted from mean RTs with right hand (SNARC index) as a function of stimulus pair for the blocked and randomized conditions (after combining the data over instruction) in Panel A and for each instruction in Panel B (after combining the data over condition), as a function of stimulus pair (plotted at the maximum of the two numbers compared) in Experiment1.

Thus, taken together, these finding replicate and extend the SNARC effects obtained by Shaki and Petrusic (2005). Moreover, these effects are clear and robust and do not depend on the direction of the instruction, nor how the instructions are presented; i.e., either varying randomly from trial to trial or fixed over a block of trials. As such, these findings serve as a benchmark for our examination of potential SNARC effects with symbolic and perceptual continua.

Experiment 2

Our knowledge concerning the perceptual properties of a wide variety of objects and events we encounter daily is maintained in long-term memory. Although considerable effort has been devoted to determining if the activated memory representations are merely at the level of an ordinal scale or whether higher order metric properties are also activated (see Petrusic & Baranski, 2002, for an overview of some of the issues), it remains to be seen if whether the mental representations activated by the symbols might also have spatial components comparable to those evident with numbers and linear orderings.

Method

Participants. Twenty-four Carleton University students participated in one 80-minute session to satisfy course requirements. All subjects reported normal or corrected-to normal vision.

Stimuli and Design. The stimulus set included three relatively small animal pairs (snail-mouse, mouse-snake, snake-goose), and three relatively large animal pairs (tiger-zebra, zebra-moose, moose-whale). The factorial combination arising from the 6 stimulus pairs, two presentation orders and two instructions was replicated ten times, preceded by three replications of practice trials.

Procedure. Participants were tested individually in a dimly lit room, seated approximately 80 cm from the centre of the video monitor. Participants were told that the presentation of either the word “SMALLER” or “LARGER” served as a warning for the next trial and was an instruction that indicated whether they were to choose either the larger or the smaller animal in the pair. After an additional 750 ms, the pair of animal names appeared while the comparative instruction remained on the screen. The participant’s task was to press the marked keyboard key (‘A’ on the left, and ‘L’ on the right) on the same side of the larger (or smaller, respectively) member of the pair. The presentation of the stimuli and the comparative instruction were response-terminated. The next trial began 1000 ms later.

The pairs of stimuli appeared at the respective centres of the left and right hemi-fields on the white background of a 17 inch (43 cm) View Sonic video monitor and the instructions, printed in David font (30, bold), appeared at the centre of the upper third of the screen. Event sequencing, randomization of trials and instructions, recording of responses and response times was under the control of SuperLab software run on a Pentium III microprocessor.

Results and Discussion

As in the first experiment, an ANOVA with instruction, the six stimulus pairs, and hand of response served as within participant factors and mean RT with correct responses defined the dependent measure. In contrast to SNARC effects with number comparisons, the interaction between stimulus pair and hand of response was not statistically reliable ($F(5, 115)=2.31$, $MS(Error)=142380.8$). Rather, the three way interaction, involving stimulus pair, hand of response, and instruction was reliable ($F(5, 115)=4.11$, $MS(Error)=52647.6$). Indeed, as is evident in the plots in Figure 2, SNARC like effects, paralleling those evident with numbers, are obtained with the instruction shorter. However, reverse SNARC like effects occur with the instruction “larger”.

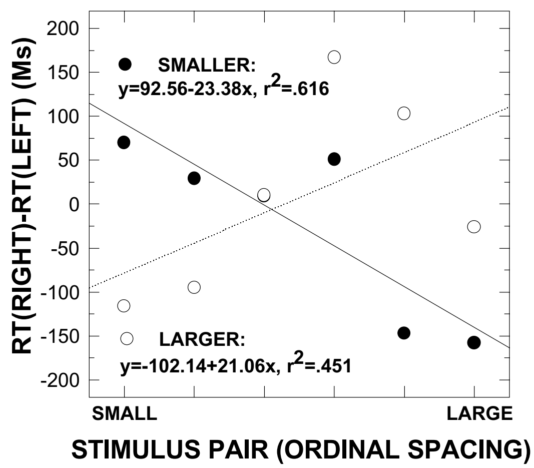


Figure 2. Mean RTs with the left hand subtracted from mean RTs with right hand (SNARC index) as a function of stimulus pair with each instruction for the animal size, symbolic comparisons of Experiment 2.

Experiment 3

Given that the mental representation of numbers is in terms of magnitude, notably extent, and spatial direction, it is natural to enquire whether the representation of visual extent parallels that of numbers. In particular, this experiment sought to determine the form of SNARC effect obtained in pairwise comparisons of visual extents.

Method

Participants. Twenty-one Carleton University students participated in two 50-min. sessions to satisfy course requirements. All participants reported normal or corrected-to-normal vision.

Apparatus. The apparatus of the two previous experiments was used except that responses were made using the buttons on an IBM-PC Mouse with the roller-ball disabled.

Stimuli. Twelve horizontal lines were used as the stimulus set. Six lines were relatively short (10, 11, 20, 21, 50, and 51 pixels) and the other six were relatively long (147, 150, 200, 210, 250, and 252 pixels). Three pairs of relatively short lines (10-11, 20-21, 50-51), and three pairs of relatively long lines (147-150, 200-210, 250-252) were created. All lines, drawn by Paintbrush software, were 1 mm wide and appeared in black on a white background. The pairs of lines appeared at the respective centres of the left and right hemi-fields on the monitor.

The two words “Longer”, “Shorter”, and the eight nonsense syllables, GUF, BIX, NIQ, YOL, ZOE, KAG, LEG, and CEB, each with 30% association value, were used as instructions in a comparative judgment of line length task. The instructions were printed in Times New Roman font (size 30, bold), and were displayed at the centre of the upper-third of the screen.

Design and procedure. Each session began with a learning phase in which the participants learned to associate each of the two instructional words (“Larger”, “Smaller”) with four of the eight CVCs. Following the learning phase, the participants were instructed that on each trial they would be presented a pair of line-lengths, and either an instructional word or one of the CVCs they had learned to associate with the instruction words in the former phase. The participant’s task was to press the mouse button on the side of the larger (or smaller) member of the pair of lines, according to the presented instruction.

Each of the six stimulus pairs, two spatial arrangements of each pair, and sixteen instructions was presented twice, for a total of 384 experimental trials, preceded by 16 practice trials. A different randomization of the 384 experimental trials was used for each participant.

Results and discussion

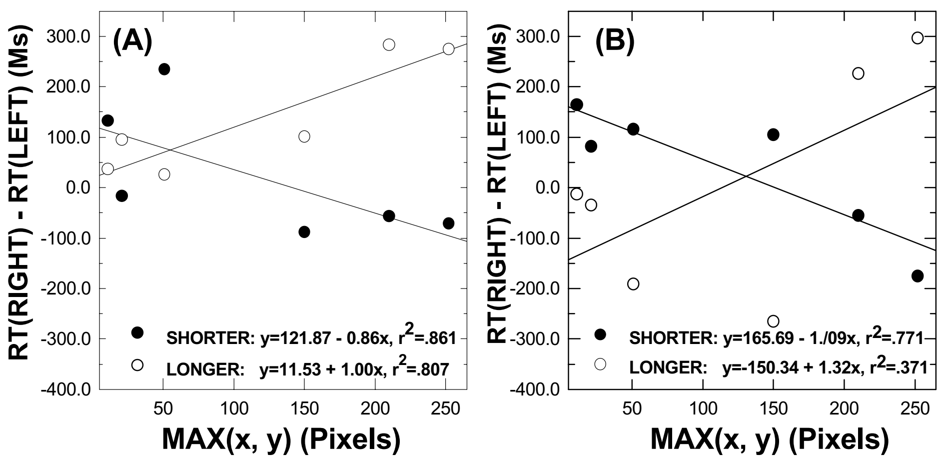


Figure 3. Mean RTs with the left hand subtracted from mean RTs with right hand (SNARC index) as a function of stimulus pair with each instruction for the line-length comparisons in Experiment 3 with the usual instructions (Panel A) and the remembered (CVC) instructions (Panel B).

As is evident in Figure 3, for both the usual instructions and the CVC instruction conditions, the interaction between stimulus pair and direction of the response is largely defined by the negative slope of the SNARC index-stimulus pair relationship with the instruction “shorter” and the positive slope with the instruction “longer”. This is then reflected in a significant linear by linear by linear component ($F(1, 18)=5.68$, $MS(Error)=257408.2$) of the three way interaction involving stimulus pair, instruction, and direction of the response.

Thus, these analyses establish reliable SNARC like effects with the instruction to choose the shorter line and a reverse SNARC like effect with the instruction to choose the longer line. Moreover, these effects are obtained with the usual instructions and when the instructions are represented in memory by nonsense syllables. Furthermore these SNARC like effects are obtained when the response codes are not tied to hand but merely to the leftward or rightward direction of the comparative response.

Summary and Conclusions

The findings of Experiment 1 replicate and extend nicely the work of Shaki and Petrusic (2005). First, as in Shaki and Petrusic, reliable and robust SNARC effects were obtained for comparisons of the magnitudes of positive number pairs. Second, also replicating Shaki and

Petrusic, identical SNARC effects were obtained with both the instruction to choose the smaller number in the pair and the instruction to choose the larger number. Shaki and Petrusic, presented the instructions in separate blocks. In the present experiment, identical SNARC effects are obtained when the instructions are blocked and when they are intermixed, thereby extending the generality of SNARC effects evident with numerical magnitude comparisons.

The present experiments also extend the SNARC effect to include spatial components in the mental representation of symbolic stimuli with underlying perceptual referents, where the association between symbol and percept is acquired naturally through experience, as in Experiment 2. In striking contrast to SNARC effects with numbers (both positive and negative), the direction of the spatial component in the mental representation is dependent on the direction of the comparison specified by the instruction. In particular in Experiment 2, with the instruction "smaller", the memory representation has the relatively small animals located on the left of the continuum and the relatively large animals on the right. On the other hand, with the instruction "larger", the memory representation activated places the relatively large animals on the left and the relatively small animals on the right. It thus appears that the dimension flows rightward from a pole defined by the presented instruction.

The mental representations of stimuli on perceptual continua also have well defined spatial components. Surprisingly, their properties parallel those evident with symbolic stimuli as is clear in Experiment 3 with comparisons of the length of horizontal line segments. When selection of the shorter line is required the mental representation activated places the short lines on the left of the continuum and the long lines on the right. However, with the instruction, to select the longer line in the pair, the relatively long lines are placed on the left end of the dimension.

It remains an important theoretical challenge to account for why the spatial representations of stimuli on symbolic and perceptual continua are dependent on the direction of the comparison while the spatial representations activated in numerical comparisons are resistant to the robust effects of the direction of the comparison. To be sure, our encounters with numerical magnitude are strongly reinforced with visual representation of a number line extending to the left and right of zero, especially in the early acquisition in school of number concepts. As well, the left-right mental number line is supported by the left right organization of numbers on keyboards and, of course, the direction of eye movements in reading. The same cannot be said of our experiences with the size of objects in memory. Rarely, are the elements in the perceptual array, of animal size, for example, laid out in a left to right array, for visual inspection. Consequently, this lack of direct experience of an ordered, left to right, array of objects varying in size, may render such mental representations more *fragile* and thus, more susceptible to the directional influences of the instruction.

References

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General, 122,* 371-396.

Gevers, W., Reynvoet, B., and Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. C*ognition, 87,* B87-B95.

Petrusic, W.M. & Baranski, J.V. (2002). Mental imagery in memory psychophysics. *Behavioral and Brain Sciences*, *25*, 206-207.

Shaki, S. & Petrusic, W. M. (2005). On the mental representation of negative numbers: Context dependent SNARC effects with comparative judgments. *Psychonomic Bulletin & Review.* (In Press).

Acknowledgement

This work was supported by a Natural Sciences and Engineering Research Council of Canada Discovery grant to William M. Petrusic

RECOGNITION AND IDENTIFICATION OF TALKERS FROM AUDITORY AND AUDIOVISUAL STIMULI

Sonya Sheffert[1], Elizabeth Olson[2] and Larence Becker[2]
[1] *Psychology Department, Central Michigan University, Mt. Pleasant Michigan 48859, USA*
[2] *Psychology Department, Towson University, Towson Maryland 21215, USA*
sonya.sheffert@cmich.edu

Abstract

The current study sought to determine if multimodal speech affects voice learning and whether these effects are moderated by variables such as the amount of prior experience with the talker, and the sensory match between learning and test context. Speaker recognition and voice discrimination were assessed following three different perceptual learning conditions: auditory-alone, audiovisual, or a combination of audiovisual and auditory exposures. Voice learning was measured using a new set of auditory-only words spoken by familiar talkers and new talkers. Listeners in both audiovisual training conditions required fewer training session to reach criteria on the talker learning task, and were more accurate discriminating known voices from unknown voices. The benefit of audiovisual training was greatest if the AV training included practice identifying the talkers from vocal cues alone. In contrast, voice learning and subsequent discrimination was poorest in the auditory-only condition. The results support the hypothesis is that visible speech gestures provide additional information about a talker's idiosyncratic speaking style and that these features are compatible with auditory talker-specific features.

.

Everyday perception relies on information from multiple modalities. In the case of speech, watching a talker's face can dramatically improve speech perception. Even under ideal listening conditions, people automatically and unconsciously combine information from both modalities (McGurk & MacDonald, 1976). These observations are consistent with the prevailing view that speech is a multimodal object of perception (Fowler, 1986, Liberman & Mattingly, 1985, Summerfield, 1987). Perceivers can recover speech gestures from multiple modes - by ear in listening, by eye in lip-reading, and even by touch in the Tadoma method of speech reading – because each mode provides information about a common articulatory event.

In addition to linguistic information, speech also conveys socially relevant information about the individual's identity and other personal qualities. Traditionally, theorists have assumed that the features of the speech signal (whether heard or seen) that carry the linguistic message are independent of the features that carry a talker's identity. And until recently, the literature on speaker recognition was devoted almost exclusively to cues which reflect a talker's unique anatomical properties, but which are phonetically irrelevant. Although qualitative, anatomically-based cues (i.e., pitch, nasality, creakiness, etc.) are important for speaker recognition, they are not the only attributes used by listeners. Idiosyncratic pronunciation habits, or "idiolect", can also be use to identify individual talkers, even those who have similar vocal tract size and shape, and who share the same regional dialect (Labov, 1986). The viability of this "ideolectical hypothesis" (cf. Remez et al., 1997), has been confirmed in the auditory domain using sinewave speech (Fellowes et al., 1997; Sheffert et al, 2003), and in the visual domain using point-light speech (Rosenblum et al., 2004). The current study builds on this work using naturalistic audiovisual speech.

The current study is based on a recent perceptual learning study by Sheffert and Olson (2004). Using videotapes of several talkers producing words, they found that listeners were more accurate at identifying familiar voices if the voices were learned in the context of the corresponding face (AV condition), despite the fact that the faces were never present during the voice identification test. However, our findings did not agree with other reports showing either negative effects or null effects of face information on voice encoding (Cook & Wilding, 1997, 2001; Legge et al., 1984; Yarmey, 1993). For example, Cook and Wilding found that a familiar voice was less likely to be recognized if it was originally encoded in the context of a videotape of the talker's face. They refer to this as a "face overshadowing effect" (FOE). They attribute the FOE to greater salience of face information when the goal is person identification, whereas vocal cues are more salient when the goal is to perceive speech. Although there are many procedural differences between their studies and our, we suspect that the amount of exposure to the talkers prior to the voice recognition test is a major reason for the discrepancies in results. Specifically, Cook and Wilding's listeners were given fewer opportunities to familiarize themselves with the talkers.

This study directly examines whether the training procedures used in Sheffert and Olson (2004) eliminate the face overshadowing effect. A new AV condition was added which might decrease voice memory relative to the previous conditions. The condition was AV training condition paired with AV generalization, rather than an auditory generalization test; this equalizes the amount of exposure to the faces and voices prior to the voice recognition test. Thus, participants received one of three training conditions: 1) AV training and AV generalization ("AV-only"), 2) AV training with A generalization ("AV-A"), and 3) A-only. Note that the latter two conditions are analogous to the AV and A training conditions used in Experiment 1. Our measure of voice learning is different, too. In Sheffert and Olsen (2004), "voice learning" was operationalized as performance on an auditory generalization task that followed each training session. It measured recognition of the five training voices; it did not require subjects to *discriminate* those five voices from five completely unknown voices. The current study was measures listeners' ability to discriminate familiar voices from unfamiliar voices following AV and Auditory-only training.

Method

Participants.

Participants were 24 adult volunteers from Towson University. All were native speakers of English with normal or corrected hearing and vision, and were paid for participating. They were randomly assigned to one of three training-generalization conditions: Auditory-only (A), AV-only (AV) and AV-Auditory (AV-A) conditions.

Stimulus Materials.

The stimulus materials consist of dynamic full-motion audiovisual tokens of ten talkers (five males and five females) producing individual single-syllable words. In all the tokens, talkers are facing the camera uttering a word in a clear, natural, emotionally neutral manner. The randomized test orders for the various training, generalization and test conditions were generated on a Macintosh computer using digital editing equipment and software (iMovie2, Macromedia Director; see Sheffert & Olson, 2004, for details). The audio track was identical for the auditory and audiovisual conditions. To control for possible differences in the "learnability" of particular talkers, four different subsets of five talkers were randomly selected; each subset included both male and female talkers.

Procedure.

Listeners were trained over several days to identify by name five speakers from the auditory or the audiovisual words. Participants were presented with a random ordering of five repetitions of ten words from each of five talkers (250 words total). Each talker spoke the same ten words within a given training session. Participants were asked to carefully listen (or listen and watch) during each trial and attempt to identify the voice of the talker. Participants were given a list of the talkers' names. After the talker had spoken and the participant had said a name, the experimenter provided the correct name of the talker. Learning criterion was 70%.

At the end of each training session, subjects completed a generalization test to determine if their talker-specific knowledge transferred to a novel set of auditory words. The generalization tests were identical to a training session, except that there was no feedback. For the Auditory and AV participants, the generalization test always matched the training condition. [This was done to maximize the likelihood of finding evidence supporting the FOE of Cook and Wilding (1997; 2001).] Participants in the AV-A condition received AV training and auditory-only generalization. This latter condition was added in an attempt to replicate and extend Sheffert and Olson (2004). Each participant who failed to achieve an average of 70% correct voice-recognition performance on the learning assessment test returned within 24 hours for another training session until he or she reached criterion. Each training-generalization session took approximately 1 hour.

After all participants reached criteria, they returned to the laboratory the next day for the voice recognition test. Before beginning the recognition test, we reassessed listeners' talker specific knowledge to confirm that they were still highly familiar with the talkers at the time of the word recognition test. To this end, participants completed the brief talker familiarization task whereby a few spoken words were aired with their associated name. Corrective feedback was not given. This task took approximately 7 minutes.

The voice recognition test consisted of 160 individual, non-repeating words presented by ten talkers. Half the words were spoken by the five talkers from the training phase, and the remaining words were spoken by five new, unfamiliar talkers. Each talker spoke an equal number of words. The words used for these lists were not used during any earlier phase of the study. Four different versions of the voice recognition list were created, each representing a different random assignment of words to talkers and a different random word order. Words were presented with a 5-s ISI, and repetitions were always identical tokens. Voice recognition was assessed by asking participants to listen to each word and decide if the voice was "old" or "new". Specifically, listeners were required to identify familiar talkers by name or, in the absence of familiarity, to respond "new". Responses were made orally and recorded by the experimenter. The voice recognition test took approximately 25 minutes.

Results and Discussion

Training and Generalization Performance.

As expected, the fastest learning occurred in the AV-only training condition (e.g., face and voice during training and generalization testing): All participants learned to associate the five audiovisual talkers with their corresponding names in a single session. In contrast, learning to recognize the talkers from auditory information alone (A-only condition) was significantly slower ($\underline{p} < .05$). On average, these participants required 2.9 sessions to learn the voice-name associations. Intermediate learning rates were observed in the AV-A condition (e.g., face and voice during training, voice only during generalization). The AV-A group took on average 2.2 days to learn the talkers, which was significantly slower than the AV-only group but significantly faster than the A-only group.

The pattern of faster voice learning in the AV-A condition relative to the A-only condition fits perfectly with the voice learning data from Sheffert & Olson (2004), indicating that our initial finding of better voice learning when visible speaker information is present during training - which had not yet been documented in the literature - was neither spurious nor limited to a particular laboratory setting. The effect replicates. Moreover, the effect generalizes to a different testing context, different sample population and different research lab.

Voice Recognition Performance

Next we asked if visible speaker information improved the ability to discriminate familiar talkers from unfamiliar talkers. All measures suggest it does. The data from the 160-item voice recognition test showed that the proportion of familiar voices correctly identified as such – that is, the hit rate – differed significantly among the three conditions, $F(2, 22) = 43.04$, $\underline{p} < .0001$, Mse = .003. The hit rate was highest in the two conditions that presented a face during talker learning (AV-only = 80%; AV-A = 89%), and considerably lower in the A-only condition (63%). Post-hoc tests indicate that each condition differed from one another (all p's < .01). The table also shows that the AV-A group had the lowest false alarm (FA) rate relative to the two other groups. However, these FA differences were not significant ($F > 1$).

To assess subjects' ability to discriminate familiar voice from and unfamiliar voices, d-prime (*d*') was computed from the proportion of correct new/old voice recognition responses and false alarms. The *d'* data are displayed in Table 1. Discrimination ability was highest among the AV-A participants, somewhat lower among the AV participants, and lowest for the A-only participants. The statistical analyses showed these differences to be significant overall, $F(2, 22) = 9.50$, $\underline{p} < .001$, Mse = .285, and to differ reliable among the three conditions (all post-hoc comparisons were significant, alpha = .05).

Table 1. Mean Proportion Correct Voice Recognition (Hit Rate), False Alarm Rate and *d'* as a Function of Training Condition

Voice Recognition Accuracy			
Training – Generalization Mode	**P(C)**	**FA**	***d'***
Auditory – only	.63	.41	.59
AV - only	.80	.41	1.13
AV - Auditory	.89	.31	1.81
Experiment Totals	.77	.37	1.18

Note. P(C) = Proportion correct "old" responses (hits). FA = False alarm: Proportion of incorrect "old" responses. *d'* = d prime.

One could argue that the high FA rate among the AV-only participants occurred because this group had the least amount of exposure to the target voices, generally only one training session before the voice recognition test. The data do not, however, support the corollary to this finding, that the group with the most exposure to the voices would have the lowest number of misidentifications. Rather, the lowest misidentification rates were found in the AV-A group, the group with target voice exposure rates between the A and AV groups. Neither is this result completely explained as a consequence of mismatched study-test context. In the AV-only condition, the learning context was perceptually and procedurally different than the test context. Although context surely plays a role in the mechanisms that underlie recognition, context alone cannot explain the pattern of recognition in the studies described here. For instance, the group with the most variable context had with the lowest misidentification rate and highest voice discrimination, whereas the group with the most consistent context (A-only) had more FAs, fewer hits and the lowest d' score.

In some respects, the data correspond well with Cook and Wilding's (2001) finding that additional exposure to a talker's face helps eliminate the tendency that the face will overshadow or interfere with voice learning. Our results indicate that the participants who had the most exposure to the AV talkers (e.g., 2 or 3 training sessions) also had, on average, higher *d'* scores on the voice recognition test.

The high voice recognition performance in the AV-A condition is possibly a result of both having experience with the face present during a portion of training, as participants with *any* AV training usually learned the voice-name association more quickly, and also from having a chance to "practice" recognizing voices from auditory materials (during the Auditory generalization test). Practice or mere exposure to the target voices in the appropriate context (i.e., auditory-alone voice test) was not the driving force behind voice identification performance, however, for participants in the A group, who had twice the amount of exposure to the appropriate context, identified the voices more poorly than did the AV-A participants. AV participants, who both saw and heard the talkers but had limited exposure, identified the voices at the lowest rate. Despite the complexity of the results, they do suggest the conclusion: Learning to identify voices is enhanced by the availability of visible speaker information, by extended exposure to the speaker, and perhaps by practice identifying talkers from auditory cues alone.

In summary, these results imply that there is common information between different modes (e.g., amodal phonetic gestures), and common information subserves different functions within a single mode (e.g., vision: lip-reading and face recognition; audition: speech and voice recognition).

References

Cook, S., & Wilding, J. (2001). Earwitness testimony: Effects of exposure and attention on the Face Overshadowing Effect. *British Journal of Psychology, 92,* 617-629.

Cook, S., & Wilding, J. (1997). Earwitness testimony 2: Voices, faces, and context. *Applied Cognitive Psychology, 11,* 527-541.

Fellowes, J. M., Remez, R. E., & Rubin, P. E. (1997). Perceiving the sex and identity of a talker without natural vocal timbre. *Perception & Psychophysics, 59*, 839-849.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics 14*, 3-28.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 816-828.

Labov, W. (1986). Sources of inherent variation in the speech process. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 402-425). Hillsdale, NJ: Erlbaum.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

McGurk, H. & MacDonald, J. W. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 651-666.

Rosenblum, L. D., Yakel, D., Baseer, N., Panchal, A., Nodarse, B., Niehus, R. (2002). Visual speech information for face recognition. *Perception & Psychophysics, 64,* 220-229.

Sheffert, S. M., Pisoni, D. B., Fellowes, J. M., & Remez, R. E. (2002). Learning to recognize talkers from natural, sinewave and reverse speech samples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1447-1469.

Sheffert, S. M., & Olson, E. A. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics, 66,* 352-362.

Summerfield, A. Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by Eye: The Psychology of Lip-reading*. (pp. 3-51). Lawrence Erlbaum Associates, Hillsdale, NJ.

Yarmey, A. D. (1986). Verbal, visual, and voice identification of rape suspect under different levels of illumination. *Journal of Applied Psychology, 71*, 363-370.

NEGATIVE MASKING, THE PEDESTAL EFFECT, AND RESPONSE CHARACTERISTICS OF THE AUDITORY SYSTEM

Daniel Shepherd and Michael Hautus
Department of Psychology, the University of Auckland, Auckland, New Zealand
daniel.shepherd@auckland.ac.nz, m.hautus@auckland.ac.nz

Abstract

A hitherto unreported relationship between the masking function and the fixed-increment function is presented and related to the units problem in audition. There are striking dissimilarities between masking functions presented in units of intensity and those presented in units of pressure. Fixed-increment functions are, however, identical regardless of which unit is used to express the level of the pedestal. By formulating a relationship between masking and fixed-increment functions it is shown that of pressure or intensity can be the correct metric of the stimulus. Data show that acoustic pressure is in fact correct. This suggests that these results can be used to address the units problem in audition and help bring a resolution to the debate.

Difference discrimination judgments require an observer to discriminate between a base stimulus (X) (known alternatively as the standard, masker or pedestal), and that base stimulus to which an increment, ΔX, has been added (i.e., $X+\Delta X$). In auditory psychophysics the masking function has long been used to test the veracity of Weber's law (i.e., $\Delta X=kX$) when observers undertake level discrimination tasks using sinusoids or noise bursts. The masking function plots the difference threshold, ΔX, as a function of pedestal level, X. When the range of X includes pedestal levels bracketing absolute threshold the masking function manifests a phenomenon known as negative masking. A masking function exhibiting negative masking is characterised by a nonmonotonic relationship between X and ΔX for circathreshold values of X. The degree of negative masking is dependent upon the choice of units representing the stimulus (i.e., the units problem). If acoustic pressure is selected as the unit then negative masking is pronounced. If units of intensity, power, or energy are employed then the magnitude of negative masking is severely diminished or eliminated.

A less commonly reported function, called by us the fixed-increment function, plots percentage correct as a function of X for the case when an increment, ΔX, is constant across all pedestal values. When values of X span the absolute threshold a phenomenon known as the pedestal effect is demonstrated. The pedestal effect consists of a nonmonotonic relationship between percentage correct and circathreshold values of X. Such a relationship is at odds with the negatively accelerating monotonic function predicted by Weber's law. The pedestal effect is observed regardless of the units representing X.

The psychometric function affords itself as a means to generate both masking functions and fixed-increment functions without having to employ separate experimental procedures. One variation of the psychometric function, the discriminability function, plots percentage correct as a function of $\Delta X/X$, where X, the pedestal, is always fixed in value. If a number of psychometric functions are generated, for different values of X, and if each of those psychometric functions shares one common absolute value of ΔX amongst them, then both masking functions and fixed-increment functions can be constructed.

Discrimination data and theoretical considerations (e.g., Laming, 1986; Pfafflin & Mathews, 1962) point to the normal probability density function as being an adequate model for the sensory effect of a stimulus, be it X or $X+\Delta X$. From this it follows that the perceived differences between two stimuli (i.e., $(X+\Delta X)-X$) can also be represented by the normal probability density function. These assumptions indicate that discriminability data can be sufficiently modelled by the upper half of a normal probability integral with respect to the difference between the two stimuli being discriminated. The psychometric function for a 2-AFC sinusoid level discrimination task is

$$P(c) = \Phi\left(\frac{\Delta X}{X} / \sqrt{2}\right) \tag{1}$$

where Φ is the cumulative normal probability density function and $\sqrt{2}$ is necessary because the variance of the distribution of differences is twice that of the original distributions. Some models of auditory level discrimination (e.g., the energy detection model) predict that the shape of the psychometric function is invariant across values of X (Green and Swets, 1966). The two psychometric functions presented in Figure 1 are based on Equation 1 and represent two arbitrary chosen pedestal values (X=40, X=70). It is demonstrated that a single psychometric function can generate both a value of the difference threshold (by putting *P(c)* equal to 0.75 in Equation 1) and a value of percentage correct for a fixed-increment (by substituting the value of the fixed-increment in place of ΔX in Equation 1). By employing a sufficient number of pedestal values it should therefore be possible to generate complete masking functions and fixed-increment functions.

The relationship between the masking function and the fixed-increment function demonstrated by Figure 1 implies values from one function can be used to predict values from the other. The shape of the fixed-increment function is invariant regardless of the units used to express pedestal level, and so should predict identical masking functions for stimuli expressed in units of pressure or intensity. However, a number of studies (e.g., Hanna, von Gierke, and Green, 1986; Viemeister and Bacon, 1988) have demonstrated pronounced differences between masking functions employing units of pressure and intensity. Thus the empirical masking function that is consistent with one predicted from an empirical fixed-increment function (and vice-versa) is likely to reflect the units of the stimulus that the auditory system operates on when analysing level. This suggests that the relationship between the two functions can be used to determine the correct units in which to express the stimulus, and hence address the units problem in audition.

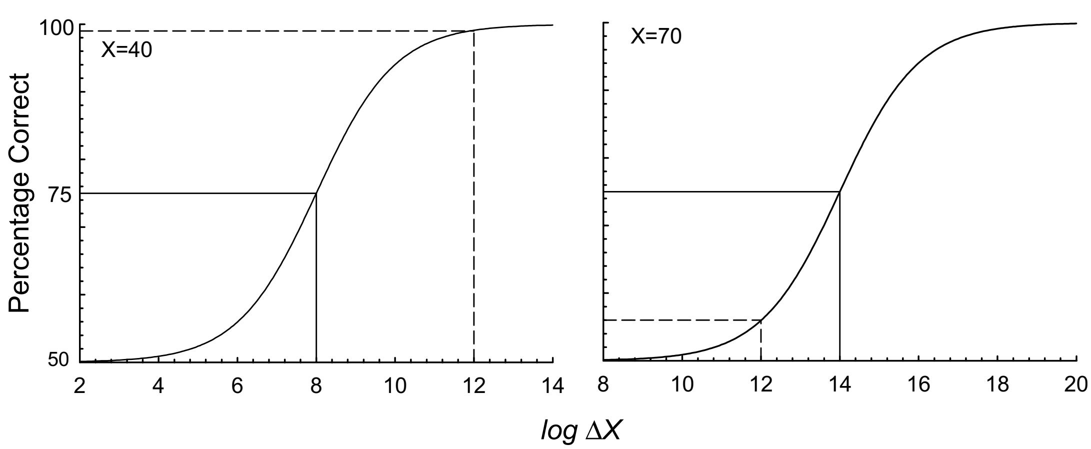


Figure 1 : Psychometric functions plotting percentage correct as a function of increment size, ΔX, for two arbitrary chosen pedestal values ($\log \cdot X$=40, $\log \cdot X$=70). The solid curve, a theoretical psychometric function, is the integral of the normal probability density function. The solid line represents the difference threshold and always intercepts the y-axis at 75 percent correct. The dashed line gives percentage correct when ΔX is fixed to 12.

Experiment 1

Method

Observers

Four experienced psychophysical observers participated in Experiment 1: DS (male aged 29), MH (male aged 36), IK (male aged 40), and BM (female aged 24). All were members of the Department of Psychology at the University of Auckland, and had normal hearing.

Stimuli and Apparatus

Stimuli consisted of 1000-Hz sinusoids of 10-ms duration. The sinusoids were constructed digitally using LabVIEW (v6.0i) and passed through a Hanning ($\cos^2$) window, with 1-ms onset and offset ramps. A National Instruments PCI-6052E card then undertook a digital-to-analogue conversion. The resulting analogue waveform was channelled to a National Instrument's BNC-2090 shielded BNC adapter chassis, and from there directed to a pair of Tucker-Davis Technologies (TDT) System 3 Programmable Attenuators (PA5). Together, the two attenuators reduced the amplitude of the sinusoids to within a safe audible range and accurately controlled their levels to those required. Once attenuated, the sinusoid was passed to a headphone buffer (TDT HB7) before being delivered to a single earphone (Telephonics, TDH-49P cradled in an MX41/AR cushion).

Procedure

Difference thresholds were estimated using a three-down, one-up adaptive procedure, yielding thresholds equal to 79% correct. Each trial consisted of two observation intervals, each containing a pedestal whose level was invariant across a block of trials. An increment was added to one of these pedestals and the observer's task was to indicate which of the two intervals contained the increment. All observers listened monaurally (left ear). There were ten levels of pedestal: -9, -6, -3, 0, 3, 6, 9, 20, 40, and 60 dB SL (re: the observer's absolute threshold). The level of the pedestal on any particular block of trials was determined randomly. After 15 reversals had occurred, the block of trials terminated. On average each block lasted 92 trials. The observer's threshold was the average of all but the first three reversal levels, which were discarded. Six such thresholds were averaged to determine the overall difference threshold for a particular level of pedestal.

Results

All difference thresholds and pedestal levels were standardised to the absolute threshold; that is, sensation level (SL) and presented both in terms of pressure ($\Delta p/p_0$) and intensity ($\Delta I/I_o$). Mean values of $20\log(\Delta p/p_o)$ and $10\log(\Delta I/I_o)$, plotted as a function of pedestal level, are plotted in Figures 2 and 3 as open squares. The zero points on both the right-hand ordinate and the abscissa represent absolute threshold. Estimates of Δp (Figure 2) for pedestals equal to or below the absolute threshold have negative values and manifest the phenomenon of negative masking. Figure 2 shows that, in this circathreshold region and in units of pressure, difference thresholds are less than absolute thresholds. In some cases the difference is greater than 9 dB SL. Beyond absolute threshold the data become linear and conform to Weber's law. The effect of expressing the discrimination data in units of intensity is demonstrated in Figure 3. Even in these coordinates there is still 2-3 dB of negative masking. A linear increment in ΔI is apparent for pedestals beyond absolute threshold. These results are consistent with the findings of Hanna, von Gierke, and Green (1986), and Viemeister and Bacon (1988).

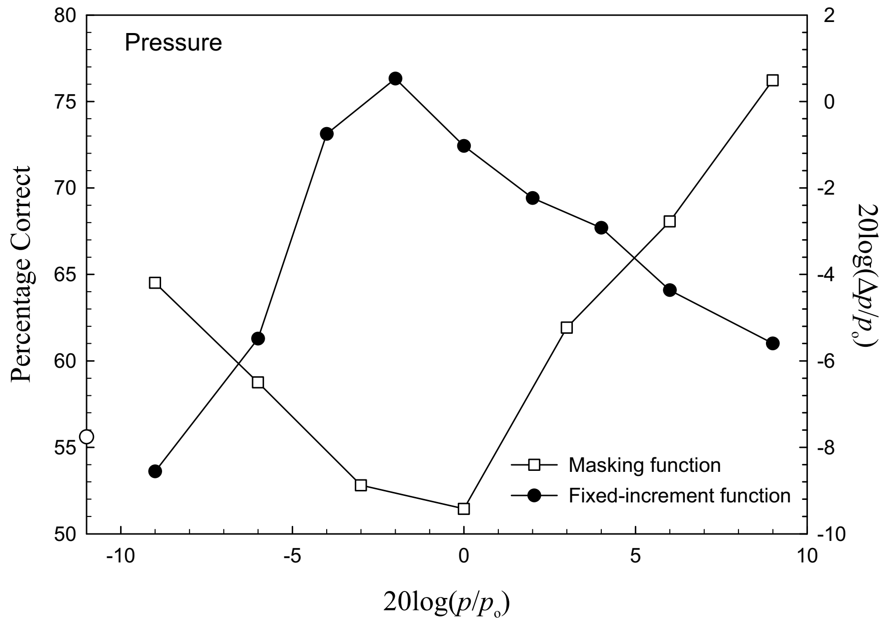


Figure 2 : Percentage correct (left ordinate) and difference thresholds (right ordinate) as a function of pedestal level, in dB SL. The fixed-increment functions (refer left ordinate) and masking functions (refer right ordinate) are expressed in units of pressure. Data consist of group means taken from Experiments 1 and 2.

Experiment 2

Method

Observers

Three of the observers (IK, BM, & DS) that had completed Experiment 1 and one other (MK, male aged 29) participated in Experiment 2.

Stimuli and Apparatus

The characteristics and generation of stimuli were identical to Experiment 1.

Procedure

Experiment 2 employed a 2-AFC procedure to estimate performance in a sinusoid-discrimination task. On any one trial each observation interval contained a pedestal of equal level and frequency, though pedestal level was variable across a block of trials. To one of these pedestals an increment was added, with each interval having an equal chance of receiving the increment. The increment, ΔX, which was of a constant value (in linear units) across all pedestal levels, was 25% of the observer's absolute threshold. Each experimental session consisted of two blocks of 120 trials each, with the first three trials of each block designated practise trials. Nine pedestal levels (-9 to 9 dB SL in 2 dB steps) were presented randomly across a block of trials. In each block of trials a pedestal level was presented thirteen times (117/9=13). Each observer undertook twenty blocks of trials in total.

Results

Fixed-increment functions are presented in Figures 2 (pressure) and 3 (intensity) and are symbolised by closed circles. Each point is the mean percentage correct values of the four observers for that particular pedestal value. The fixed-increment functions presented in Figures 2 and 3 are in fact identical. The fixed-increment function peaks at the -2 dB SL pedestal and exhibits a linear decline as the pedestal level both decreases and increases from this value. The average of the no pedestal condition, in which the increment was presented in isolation, is the open circle, o, intercepting the y-axis, and is equal to 55.57%. The pedestal effect is evident in that observers perform better when an optimal subthreshold pedestal is present than if a suprathreshold pedestal or no pedestal is added. Note that the level of pedestal associated with the highest percentage correct is itself below the absolute threshold. This indicates that discriminability is improved by the provision of subthreshold pedestals, and mirrors the findings reported by Pfafflin and Mathews (1962).

Discussion

Figures 2 and 3 plot the masking function and the fixed-increment function on the same graph, in terms of pressure (Figure 2) and intensity (Figure 3). For each graph both types of plots share a common abscissa (viz pedestal level) while the masking function uses the right-hand scale (the difference threshold) and the fixed-increment function the left-hand scale (percentage correct).

A point will be now raised regarding the relationship between the masking function and the fixed-increment function. First, recall that each difference threshold represents 79 percent correct in a 2-AFC three-down, one-up adaptive procedure. Consider the hypothetical case in which, across a range of pedestal values, the difference thresholds, ΔX, are all equal. In such a scenario the masking function would be best modelled by a straight horizontal line. If this value of ΔX was retained, along

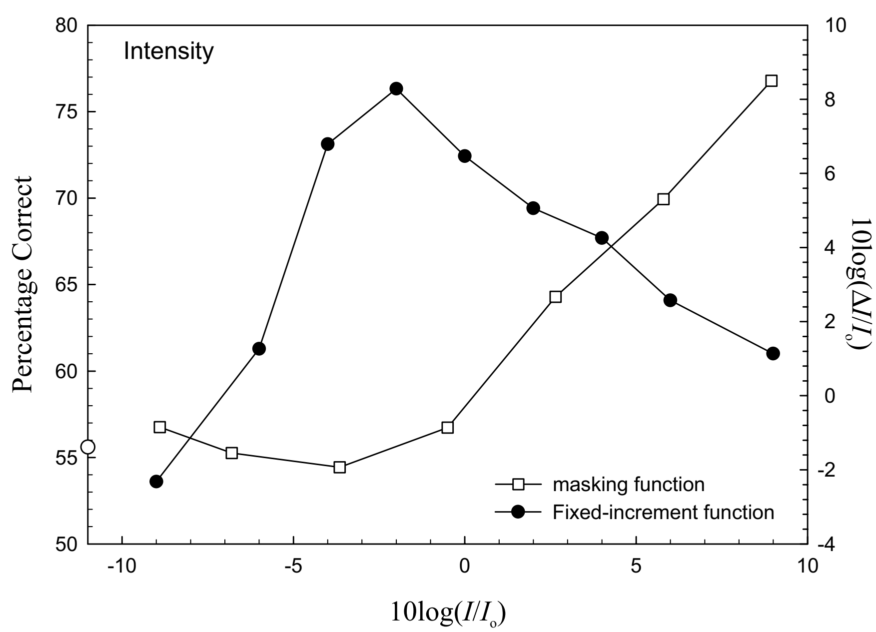


Figure 3 : Percentage correct (left ordinate) and the difference threshold (right ordinate) as a function of pedestal level, in dB SL. The fixed-increment functions (refer left ordinate) and masking functions (refer right ordinate) are expressed in units of intensity. Data consist of group means taken from Experiments 1 and 2.

with the same range of pedestals, and used in a 2-AFC discrimination task it would be expected that at each level of pedestal the performance of the observer would be approximately 79%. This latter data, percentage correct as a function of pedestal level for a fixed value of ΔX, is of course the fixed-increment function.

Now consider the case of a masking function possessing a set of difference thresholds that are not equal and the subsequent effect of these inequalities on the fixed-increment function. Smaller values of ΔX, where X could be either p (pressure) or I (intensity), should, for the same pedestal values, correspond to a higher percentage correct score on the fixed-increment function than larger values of ΔX. This is because, at pedestal values associated with small values of ΔX, the auditory system is more sensitive to the addition of an increment than pedestals associated with higher values of ΔX. This difference in sensitivity will be reflected in the percentage correct values of the fixed-increment function. Thus an inverse relationship of sorts is expected between the masking and the fixed-increment function.

When expressed in units of pressure (Figure 2), the relationship between the fixed-increment function and the masking function appears to uphold the inverse relationship. Figure 3 shows that the first four difference thresholds are approximately equal when intensity is the unit of choice. It would therefore be expected that across this range of pedestals the percentage correct values reflected in the fixed increment function would also be equal to one another. It is clear that they are not. On the basis of this evidence it appears that pressure produces more meaningful measurements than intensity.

References

Green, D.M., and Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hanna, T.E., von Gierke, S.M., and Green, D.M. (1986). Detection and intensity discrimination of a sinusoid. *The Journal of the Acoustical Society of America*, 80, 1335-1340.

Laming, D. (1986). *Sensory Analysis*. London: Academic Press.

Pfafflin S.M., and Mathews, M.V. (1962). Energy-detection model for monaural auditory detection. *The Journal of the Acoustical Society of America*, 34, 1842-1853.

Viemeister, N.F., and Bacon, S.P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *The Journal of the Acoustical Society of America*, 84(1), 172-178.

EXPECTANCY, NOT MEMORY DETERMINES IDENTICAL SEARCH RATES IN STATIC AND DYNAMIC DISPLAYS

[1]Zhuanghua Shi, [1,2]Mark A. Elliott
[1]Department Psychologie,
Ludwig-Maximilians Universität, Munich/München, Germany
[2]Department of Psychology
National University of Ireland, Galway, Ireland

Abstract

Search rates are often used as a measure of search efficiency. Horowitz & Wolfe employed this measure to compare static and dynamic visual search (Horowitz & Wolfe, 1998). Based on identical search rates, they argued that visual search operates without memory. Subsequently and in contrast a number of other authors have agued that visual search relies upon iconic memory (Bäcker & Peral, 1999; Peterson, Kramer, Wang, Irwin, & McCarley, 2001; Scheier, Khurana, Itti, & Koch, 1999). In a reanalysis of Horowitz and Wolfe's data, Kornbrot found contrary results (Kornbrot, 2004). Search rates only indicate search efficiency and the implications of this may be limited. In this commentary we propose a probability model and point out that factors such as target probability and subjective expectancy may bring about similar results.

In a typical visual search task, subjects are asked to search for a target positioned amongst distracters, for example for a red diagonal bar nested among green vertical bars. In the majority of experiments there are two common dependent variables, response time (RT) and accuracy, which are employed as indexes of search efficiency and by extension as indicators of the mechanisms of search. More specifically, search rates (or search slopes), i.e. RT increment per item, are a particular focus of interest (Wolfe, 1998) and have lead to concepts such as parallel/serial search (Treisman & Gelade, 1980) and guided search (Wolfe, 1994). A typical serial search model assumes that search rates reflect some degree of sequential processing. Once an item has been processed, it is tagged and never again reprocessed. Implicit in this assumption is the notion of memory-driven search. This assumption was questioned by Horowitz and Wolfe (1998), who used static and random displays for the presence of a target. In random display, all item locations were changed every 111 ms, whereas in static display all item locations remained the same. Based on identical search rates and the impossibility of a memory-driven strategy in random displays they argued visual search operates without memory. This was soon subject to criticism by a number of researchers citing eye movement data (Peterson et al., 2001), the inadequacy of Horowitz and Wolfe's data (Bäcker & Peral, 1999), and more recently reanalysis of their data (Kornbrot, 2004) as indicating quite the contrary. The counter claim being that visual search does have a memory.

In spite of this debate for a number of years Townsend and others have repeatedly warned against using the method of varying set size and measuring response time for differentiating serial and parallel search (Atkinson, Holmgren, & Juola, 1969; Townsend, 1971, 1990, 2001). In current paper we show that even adopting a strict serial processing assumption it is still dangerous to conclude anything about the role of memory in visual search from search rates alone.

Probability models in visual search

Horowitz employed two different search displays (Horowitz & Wolfe, 1998). One condition was a normal static search display with set size *n*, the second was a random dynamic display with same amount of items but with item locations shifted randomly every 111 ms. The search tasks were the same. For simplicity, we will describe only serial self-terminating search model. It's enough to see how search rates can be identical with memory assumption, although an equivalent parallel model can be developed. A serial self-terminating model assumes that one item is processed at a time and search is terminated after a target has been found or all of the items have been examined. Furthermore, the distribution of processing time for the individual target and distracter items was also assumed to be identical. Under these assumptions, the mean RTs in Horowitz's experiments for static and random display are equivalent to average checked balls in following two classical probability questions, respectively:

1. There're n balls in a cloth bag with colour hidden from view. The balls are red or blue. With a 50% probability the bag contains one red ball and the experimental subjects are allowed to take out one ball at a time. What is the mean average number of times taken for a subject to find the red ball?

2. The same basic setting as above with a variation in procedure. Subjects are allowed to take out one ball at a time but if the ball is not red it must be returned to the bag. The question again asks, what is the mean average number of times taken for a subject to find the red ball?

Since Experiment 3 in Horowitz's paper is only considered in terms of a target being present, the same situation is also considered here. The answer to first question is:

$$\begin{aligned} Mean &= \sum_{i=1}^{n} i \cdot \Pr(step\ i\ get\ red\ ball) = \sum_{i=1}^{n} i \cdot \prod_{i<i} \Pr(ball\ \ j\ is\ blue) \cdot \Pr(ball\ i\ is\ red) \\ &= \sum_{i=1}^{n} i \cdot \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdots \frac{1}{i} = \frac{n+1}{2} \end{aligned} \tag{1}$$

For the second question, because of replacement of distractor balls, the probability of a red ball (target) in each step is identical, i.e. $p = 1/n$ while probability of blue balls (distracters) is $q = 1\text{-}1/n$. Thus:

$$\begin{aligned} Mean &= \sum_{i=1}^{M} i \cdot \prod_{i<i} \Pr(ball\ \ j\ is\ blue) \cdot \Pr(ball\ i\ is\ red) = \sum_{i=1}^{M} i \cdot q^{i-1} \cdot p \\ &= \frac{1}{p} \cdot (1 - q^{M}) - M \cdot q^{M} \\ &= n - (n + M) \cdot \left(1 - \tfrac{1}{n}\right)^{M} \end{aligned} \tag{2}$$

Where M is maximum number of repeated checks, we refer to this as a stopping criterion. In an ideal model M should approach infinity. However, because of the small probability for very late target appearance and subject impatience, subjects always set some limited criteria (Experiments showed trials with response longer than 5s were less 2% in total, Horowitz & Wolfe, 1998). Kornbrot argued that these criteria should independent of set size due to dependent error rates (Kornbrot, 2004). Bäcker and Peral (1999) made a similar assumption since they estimated M around 33 for all set sizes. Whether stopping criterion M is set size independent or not, error rates should be detailed examined. It is quite clear, without considering misses that pure error rates when applying stopping criterion M are $e = (1 - \frac{1}{n})^{M}$.

Detailed relationships between error rates and stopping criterion M are shown in Figure 1a. The real error rates from the behavioral data (the triangles in Figure 1a, see also Figure 1 in Kornbrot) indeed suggest M is not independent of set size n. This is contrary to the conclusion reached by Kornbrot.

A reasonable alternative assumption is that in dynamic search subjects set their stopping criterion M as a function of set size and with an initial intercept C, i.e.

$$M = C + k \cdot n \tag{3}$$

Under these conditions, set size determines stopping criterion. We further investigated how these two different conditions, i.e. static and random displays, can produce similar search rates without assuming the involvement of any other factors.

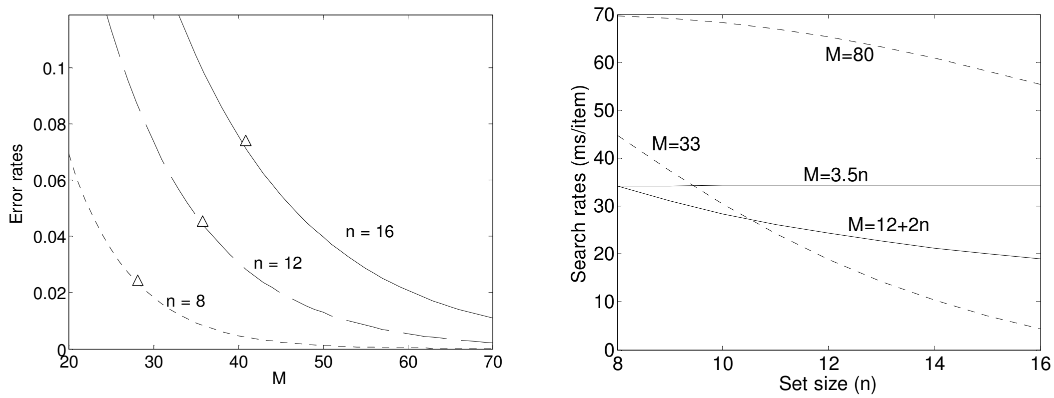


Fig. 1. a. Left panel: Error rates as functions of stopping criterion *M* (the maximum number of items checked) for different set sizes. Dot line: random display with set size 8, dash line: set size 12 and solid line: set size 16. Triangles are the empirical error data from Horowitz's Experiment 3. Right: Search rates as a function of set size for different stopping criterion *M*. dash lines: uniform stopping criteria. Solid lines: set size depended criteria.

Because of linear properties (Equation 1) the search rate of static condition is straightforward,

$$s = 0.5\tau, \tag{4}$$

where τ estimated from Horowitz's experiment 3 is 70ms and search rate s is 35ms. The search rate of dynamic condition is a little more complex:

$$s = \left(1 - (n+1+M)\left(1 - \frac{1}{n+1}\right)^M + (n+M)\left(1 - \frac{1}{n}\right)^M\right) \cdot \tau \tag{5}$$

Note this is a nonlinear function of set size. Principally the search rates are very different from the static condition.

Based on Equation 5, Figure 1b shows how different stopping criteria affect search rates as a function of set size. If subjects have patience and set a very large termination time, the search slopes should actually double relative to the static condition. For example search slope is near 70ms in random condition compared to 35ms in static condition for set size 8 and M equals 80 (Figure 1b M=80). This predicted the Horowitz's Monte Carlo simulation (Horowitz & Wolfe, 1998). However, in most cases subjects terminate search quite early. Bäcker estimated that subjects terminated search roughly after examining 33 items for random displays. Following this stopping criterion, however, Figure 1 b shows that search rates decrease dramatically when set size increases, although the mean average of 3 levels (set sizes of 8, 12 and 16 distracters) was similar to the empirical data. A possible ideal stopping criteria which can guarantee similar search rates to the static condition is an absolute set size dependent termination strategy (Figure 1b. *M=3.5n*). Under this strategy search rates are kept almost the same as the static condition among a wide range of set sizes. Furthermore the error rates should also remain as low as under static conditions (Figure 1a). However, the real data suggests that slopes decrease slightly with increasing set size (Figure 2a in Horowitz, 1998) while error rates increase as a function of set size (c.f. Figure 1a triangles). Consequently, the most probable strategy undertaken by subjects would be to relative set size dependent termination strategy (Figure 1b. *M=12+2n*). With this strategy the model predicts both Horowitz's error rates and search slopes more precisely.

Discussion

Search rates and memory

Inferring mechanisms from search rates has been counselled against for a number of years (Atkinson et al., 1969; Townsend, 1971, 2001). These commentators warned that search rates are inadequate a means to discriminate between parallel and serial search. The above probability models also show there is a similar danger in inferring whether search is memory driven or not. The first non-replacement probability model presented here has an implicit memory assumption. By contrary, there's no memory assumption in the second replacement model. Even with these two different assumptions modelled search rates can be identical by using set size dependent termination criteria (Figure 1b). Search rates merely reflect the search efficiency. On this basis one should exercise considerable caution in inferring from search rates alone, especially when examining a very few levels of set size.

Stopping criteria and expectancy

Contrary to the set size independent strategies proposed by Kornbrot and Bäcker (Baecker & Peral, 1999; Kornbrot, 2004), the above analyses indicated that stopping criteria M are dependent of set size n. In random search displays, target probability is an inverse property of set size n. Therefore, stopping criteria is actually related to target probability. A lower target probability makes for an expectancy of longer search. When set size increases and target probability decreases in random search display, subjects implicitly increase stopping criteria M, with the result that, as a by product –both static and dynamic search produce similar search rates.

Acknowledgements

This research was supported by German research council (DFG) grant EL 248/1 to MAE.

References

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. (1969). Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics, 6*, 321-326.

Bäcker, A., & Peral, E. (1999). *Visual search: amnesic or undecided?* Paper presented at the Proc. 3rd Annual Vision Research Conference, Fort Lauderdale, FL.

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature, 394*(6693), 575-577.

Kornbrot, D. E. (2004). *In the glimpse of an eye: decision making and vision.* Fechner Day 2004, Coimbra: Gráfica de Coimbra, Lda. (pp. 252-257

Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychol Sci, 12*(4), 287-292.

Scheier, C., Khurana, B., Itti, L., & Koch, C. (1999). *Visual Search Amnesic or Memory Driven?* Paper presented at the Proc. 3rd Annual Vision Research Conference, Fort Lauderdale, FL.

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics, 10*, 161-163.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46-54.

Townsend, J. T. (2001). A clarification of self-terminating versus exhaustive variances in serial and parallel models. *Percept Psychophys, 63*(6), 1101-1106.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Wolfe, J. (1998). *Visual Search*. London, UK: Universitz College London Press.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*(2), 202-238.

REPETITION PRIMING WITH BARELY DISTINCT STIMULI

Alexander Sokolov [1], and Marina Pavlova [2]
[1] ZNL Center for Neuroscience and Learning, Department of Psychiatry III, University of Ulm Medical School, Leimgrubenweg 12, D 89075 Ulm, Germany
[2] Developmental Cognitive and Social Neuroscience Unit, Children's Hospital, and Institute of Medical Psychology and Behavioral Neurobiology, MEG-Center, University of Tübingen, D 72076 Tübingen, Germany
e-mail: alexander.sokolov@uni-ulm.de

Abstract

In repetition priming studies, frequently repeated stimuli typically yield shorter response times (RT) and lower error rates compared to rarely presented stimuli. Some, however, report a reverse repetition priming effect. Here we ask whether repetition priming occurs with barely distinguishable stimuli and how, if at all, these effects depend on their presentation order. Participants had to accomplish visual binary classification task without practice and feedback, assigning one of two gray disks (size difference, 5%) into either small or large category by pressing a respective key. In a 2 x 2 design, we varied the frequency of small and large disks in the sets (3:1 or vice versa) and the serial order of their presentation (either small or large ones were more likely to occur at the series outset). The results indicate that with the average percentage correct of 69 %, the stimuli were only barely distinguishable. Unlike usual repetition priming effects, (i) for RT, repetition priming was highly stimulus- and experimental set-specific, occurring with the small disks in the concordant sets (with frequent stimuli presented mainly at the series' outset), and with the large disks in the discordant sets (with mainly infrequent stimuli occurring at the outset). With the large disks in concordant series and small disks in discordant series, RT repetition priming was reversed. (ii) Similarly, the reverse repetition priming effects were found for the error rate, with the frequently repeated - both small and large - disks giving rise to higher error rates. We conclude that in contrast to the findings obtained with well-discernible stimuli, barely distinct stimuli produce reverse repetition priming effects. Repetition priming is likely to engage multiple neural mechanisms.

Repetition priming represents either facilitatory or inhibitory changes in the brain and behavioral response to a current stimulus that has been already encountered on previous occasions. This generic class of priming phenomena is viewed as an elementary form of implicit episodic memory (e.g., Schacter, Dobbins, & Schnyer, 2004). Repetition priming can be short-term and long-term, depending on the time elapsed between the two consecutive encounters of the target stimulus. Current psychophysical and neuroimaging work strives to uncover neurobiological mechanisms of repetition priming using chiefly well-discernible stimuli (e.g., Henson 2003; Schacter et al., 2004). In these studies, frequently repeated stimuli typically yield shorter response times (RT) and lower error rates (i.e., performance facilitation) compared to rarely presented stimuli. For example, when the participant is required to judge if an object in a string [*apple*, *chair*, *pencil*, *apple*, *chair*, *apple*,..] is bigger than a shoebox, the yes/no response will be faster and more accurate to *apple* than to any other, less frequently presented object.

Some, however, report a reverse repetition priming effect, that is, longer RT and higher error rates for the frequent compared to rare stimuli (i.e., an inhibitory change in performance). *Negative* priming effects have been found in short-term priming for compound primes that comprise both attended and ignored stimuli (e.g., in a flanker task by Eriksen & Eriksen, 1974). The responses are slower and less accurate to the targets that match previously ignored stimuli (Tipper, 1985; see Fox, 1995; May, Kane, & Hasher, 1995; Neill & Mathis, 1998 for review). In a short-term masked priming task with subliminal primes, Eimer and Schlaghecken (1998) and Eimer (1999) obtained longer RT and more errors on compatible trials (with the matched prime and target) than on incompatible trials (when the prime and target did not match). Still another example represents *repetition blindness*, that is, a reduced accuracy in reporting and detecting the stimuli that have been seen shortly before (Kanwisher, 1987; Hochhaus & Johnston, 1996).

Here we ask whether long-term repetition priming occurs with barely distinct stimuli and how, if at all, these effects depend on the stimulus presentation order. In contrast with well-discernible stimuli, barely distinct ones exhibit only a small difference in sensory information available and a highly arbitrary linkage to the required response; hence the role of non-sensory variables (like, for example, response bias) in processing such kind of stimuli is greatly increased. We expected, therefore, that repetition priming with barely distinct stimuli might be different from the pattern of results commonly found with well-discernible stimuli.

Method

Participants

Four separate groups of student volunteers (20 female and seven male; mean age, 26.1 years) took part in the study for a course credit. They had normal or corrected vision, were unaware of the aim of the research, and run individually.

Stimuli, procedure, and generation of presentation series

In order to reduce the amount of and the difference in sensory information between distinct items, we used as stimuli two, small and large, gray disks that only slightly differ in their size (diameter, 6° and 6.3°; size difference, 5%). They were presented one per trial for 300 msec on an otherwise black computer screen (15”) throughout the experiment and viewed from a distance of 57 cm. The stimuli were generated by a Microsoft Windows operated PC using MatLab 6.2 and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), which also controlled stimulus delivery and recorded participants’ responses.

In a 2 x 2 factorial design, four independent groups of participants were presented with experimental sets of the stimuli that differ in their frequency skew (with the frequent small or large disks; proportions, 3:1 or vice versa) and serial presentation order (either frequent or infrequent, small or large, disks were more likely to occur at the series outset). In two groups, 200 trials (ratio of small/large disks, 150/50 or 50/150) were randomized for presentation in a standard, computer-assisted way, yielding frequent stimuli to occur mainly at the beginning of the presentation series (concordant series; positive skew and order for frequent small disks and negative skew and order for frequent large disks). By contrast, in the other two groups, the trials were randomized with a bias, such that mainly infrequent stimuli did so (discordant series; positive skew but *negative* order for frequent small disks and negative skew but *positive* order for frequent large disks). Each participant accomplished 2 similar blocks of 200 trials according to the group assignment. The experimental procedure and algorithm for generating the presentation series are described in more detail elsewhere (Sokolov, Reissner, & Pavlova, 2004).

A trial started with a fixation point in the middle of the black screen that appeared for 500 msec, followed by a disk and then by a black screen. The participant had to accomplish a visual binary discrimination task, assigning the disk into either *small* or *large* category, as rapidly and accurately (according to their own feeling) as possible pressing a respective key. The next trial started one second upon the key press. The assignment of response keys (left/right to *small/large*) was counterbalanced across participants. While judging the size, participants were required not to compare a current stimulus to the previous ones and to rely only on their own impressions. Their head movements were restricted by a chin-rest. Neither prior information about the number of distinct stimuli, their probabilities and serial order, no familiarization/practice trials or feedback of performance were given to the participants.

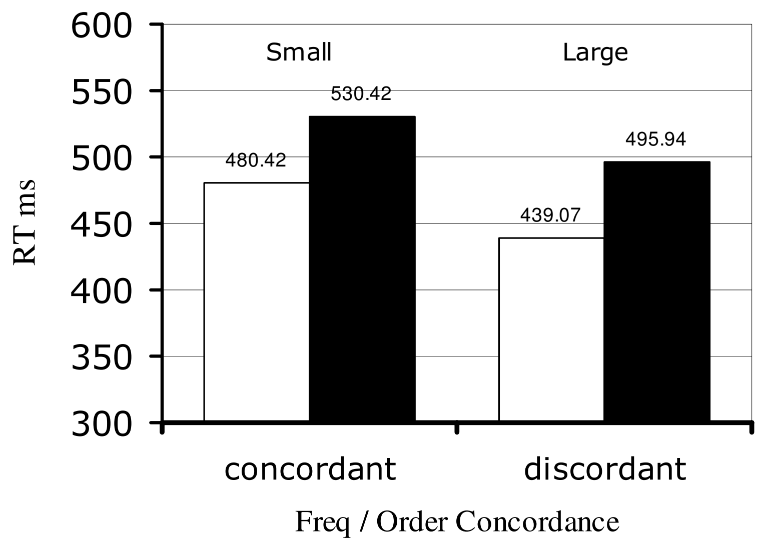


Figure 1. Mean response time (RT, ± 3SD corrected) for the frequent, *white bars*, and rare, *black bars*, - small and large - disks in the concordant and discordant experimental series. Usual repetition priming effects: faster responses occur for the frequently presented stimuli.

Results

The results indicate that with the group average percentage of 69% correct, the stimuli were only barely distinguishable. Unlike usual repetition priming effects with well-discernible stimuli, repetition priming for RT in this experiment was highly stimulus- and experimental set-specific, occurring with small disks in the concordant sets (when frequent – small or large - stimuli were presented mainly at the series' outset), and with large disks in the discordant sets (when mainly infrequent – small or large - stimuli occurred at the outset). Figures 1 and 2 show the mean RT for correct responses (±3 SD corrected values) to the frequent and rare stimuli (small and large disks separately) averaged across participants in the concordant and discordant experimental series. Figure 1(left) indicates that in concordant series, RT is considerably shorter for frequently presented than for rarely presented small disks. In the discordant series, faster responses occur for the frequent compared to rare large disks (Figure 1, right).

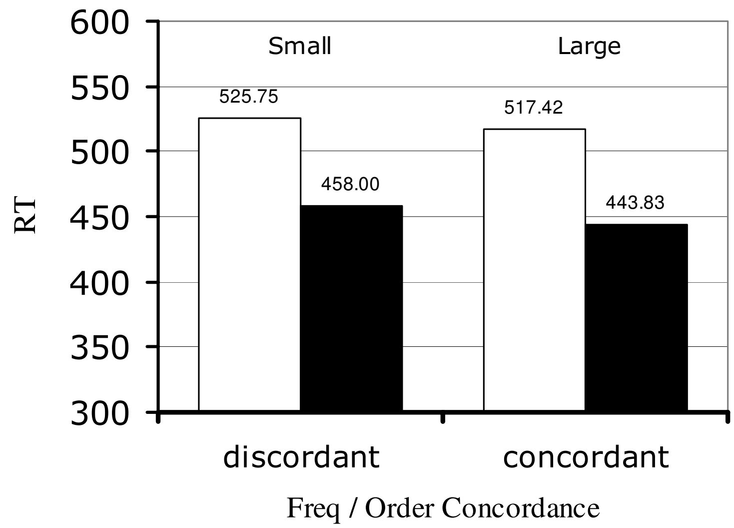


Figure 2. Mean response time (RT, ± 3SD corrected) for the frequent, *white bars*, and rare, *black bars*, - small and large - disks in the concordant and discordant experimental series. Reverse repetition priming effects: faster responses occur for the rarely presented stimuli.

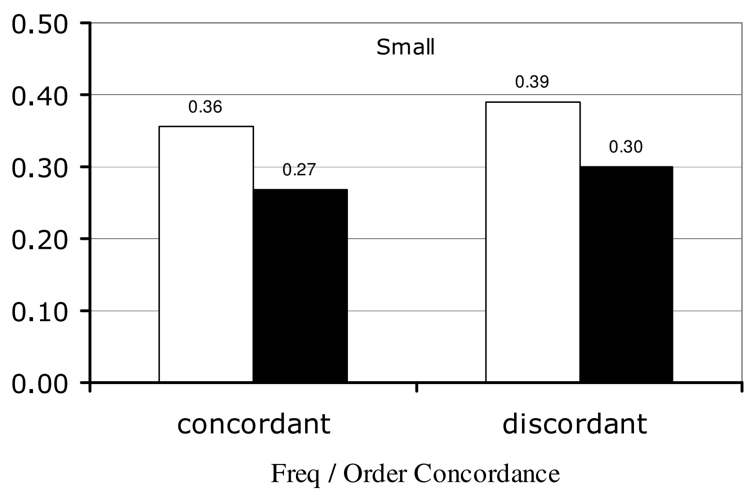


Figure 3. Mean error rates for the frequent, *white bars*, and rare, *black bars*, small disks in the concordant and discordant experimental series. Reverse repetition priming effects: fewer errors are made with the rarely presented stimuli.

By contrast, with the large disks in concordant series and small disks in discordant series, RT repetition priming was reversed. That is, RT is longer for frequent than for rare either small or large disks in the discordant and concordant series, respectively (Figure 2).

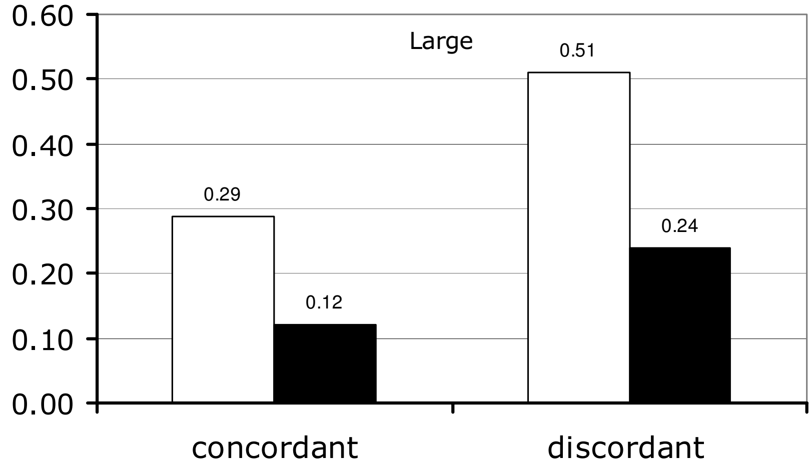


Figure 4. Mean error rates for the frequent, *white bars*, and rare, *black bars*, large disks in the concordant and discordant experimental series. Reverse repetition priming effects: fewer errors are made with the rarely presented stimuli.

Similarly, the reverse repetition priming was found for the error rates in all conditions, with the frequently repeated - both small and large - disks giving rise to higher error rates. Figures 3 and 4 show the mean error rate for the frequent and rare stimuli (again, separately for small and for large disks) averaged across participants in the concordant and discordant series. As can be seen from these Figures, the frequent stimuli yield a greater number of erroneous responses than do the rare stimuli.

Discussion

The outcome of the present study clearly indicates that in comparison with well-discernible stimuli, barely distinct stimuli produce mostly the reverse priming effects. These effects, particularly for error rates, did not seem to depend on the serial presentation order of stimuli, that is, on whether on overall frequent or infrequent stimuli dominated the outset of a presentation series (concordant and discordant series, respectively). For RT repetition priming, however, the sign of the effects (facilitation or inhibition, i.e., the reverse priming) was specific to the stimulus used and the contingency between the frequency skew of the set and the serial presentation order of stimuli from that set. Commonly observed facilitatory, faster responses to the frequently repeated stimuli occurred only with the small disks presented in concordant series and with the large disks in discordant series. With the small disks presented in discordant series and large disks in concordant series, the repetition priming was reversed, indicating an inhibitory-like slowing down of the responses to frequent stimuli.

The prevalence of reverse repetition priming effects with barely distinct, compared with well-articulated, stimuli may be accounted for either by stimulus-, stimulus-response associative, or response-related processing (e.g., Schacter et al., 2004). The difference in sensory information available in the stimuli used in the present study is negligible while the stimulus-response associations are highly ambiguous. In addition, an earlier study (Sokolov et al., 2004) indicated a substantial modulation of the response bias by the experimental conditions used in this study. Therefore, we conclude that the locus of the effects may be found either in response-related or associative, rather than stimulus-specific processing. Previous research with subliminal masked primes has also suggested that the reverse short-term priming effects can be generated by a competition of two response tendencies activated in succession by the prime and target stimuli (Eimer & Schlaghecken, 1998; Eimer, 1999). Future work is required for understanding the precise nature, psychophysical and neurobiological mechanisms of these effects.

Acknowledgements. This work was supported by a grant # SO 465/7 to A.S. from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG). We thank Timo Reissner for the assistance in conducting the study.

References

Brainard, D.H. (1997). The Psychophysics toolbox. *Spatial Vision*, **10**, 433-436.

Eimer, M. (1999). Facilitatory and inhibitory effects of masked prime stimuli on motor activation and behavioural performance. *Acta Psychologica*, **101**, 293-313.

Eimer, M., & Schlaghecken, F. (1998). Effects of masked stimuli on motor activation: Behavioural and electrophysiological evidence. *Journal of Experimental Psychology: Human Perception and Performance*, **21**, 1737-1747.

Eriksen, B.A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, **16,** 143-149.

Fox, E. (1995). Negative priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review*, **2**, 145-173.

Henson, R.N.A. (2003). Neuroimaging studies of priming. *Progress in Neurobiology*, **70**, 53-81.

Hochhaus, L., & Johnston, J.C. (1996). Perceptual repetition blindness effects. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 355-366.

Kanwisher, N. (1987). Repetition blindness: Type recognition without token identification. *Cognition*, **27**, 117-143.

May, C.P., Kane, M.J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin*, **118**, 35-54.

Neill, W.T., & Mathis, K.M. (1998). Transfer-inappropriate processing: Negative priming and related phenomena. In: D.L. Medin (Ed.), *The psychology of leaning and motivation: Advances in research and theory* (Vol. 38, pp. 1-44). San Diego: Academic Press.

Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, **10**, 437-442.

Schacter, D.L., Dobbins, I.G., & Schnyer D.M. (2004). Specificity of priming: A cognitive neuroscience perspective. *Nature Reviews Neuroscience*, **5**, 853-862.

Sokolov, A., Reissner, T., & Pavlova, M. (2004). Change the base rate but beware of order: Discriminating the visual size in variable contexts. In: Oliveira, A.M., Teixeira, M., Borges, G.F., & Ferro, M.J. (Eds.), *Fechner Day 2004* (pp. 541-546). Coimbra, Portugal: The International Society for Psychophysics.

Tipper, S.P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **37A**, 571-590.

Eighty Years & Counting: Helson's Influence on the Study of Context in Psychophysical Judgments.

Joseph Steger, Ernest M. Weiler,
David E. Sandman, Joel S. Warm
University of Cincinnati, Cincinnati, Ohio.
Contact Ernest M. Weiler Ph.D. ML #379
University of Cincinnati, Cincinnati, Ohio 45267

Abstract

The purpose of this paper is advocate the historical relevance of Harry Helson for his work and theory on the importance of context in Psychology, particularly in *Psychophysics*. We know of no one who questions the importance of context in modern research, but we are concerned that credit for Helson's historical contribution to the Zeitgeist of research is fading. Just 80 years ago, Helson published his dissertation work in an influential series of articles interpreting (from the original German) the emerging theory of Gestalt Psychology, the "principle of the whole" (Helson, 1925; 1926), adding a summary of 114 principles of Gestalt in 1933. Over the next 40 years, from that original work to his text *Adaptation Level* (1964a) he contributed a series of insightful, stimulating and even controversial experiments, on the influence of the frame of reference/context in psychophysics and in other areas of psychology (e.g. Helson, 1947; 1948, 1951, 1964b).

Helson pursued work on context at a time when the search for ***simple input-output*** relations dominated the field of psychophysics. The focus of his research was to show that such a view was and is too elementary and that sensory discrimination depends not only on the immediate stimulus but also upon the contextual background in which it appears. As his contemporary, Floyd Allport wrote in describing Helson's work in his classic volume on theories of perception (Allport, 1955), "We might well seek some formula more comprehensive than that of psychophysics…The sensory experiences studied in psychophysics are stimulus-bound...considered as somewhat standard, that is, pretty much the same for everybody,... [but] Under certain conditions individuals acquire a scale, a frame of reference, by which to judge and this scale is subjective and personal to the subject. It is upon such considerations that Helson's concept of the "adaptation level"... is based." Allport then quotes Helson as proposing " a concept not designed to replace other theories but to supplement and broaden them." In effect, as Corso (1967) pointed out, Helson sought to alert psychophysicists to the fact that factors heretofore considered as ***sources of noise*** non-essential to sensory discrimination, such as an individual's attitudes, values, and ways of structuring his experience, do not act willy-nilly on the organism, but indeed do matter in important ways. For example, Helson argued that Weber's law, Delta I/I = k might be rewritten as :

$$\text{Delta } A = k\,(X_i + A)/2$$

in which A = the adaptation level, which results from the pooling of the stimulus series, k = Weber's constant and X_i = a particular stimulus of the series being judged (Allport, 1955). The point is that the judgment of the individual stimulus will be altered by the effects of the series; there is no absolute stimulus-bound answer to the discrimination issues raised by Weber's law, or any other psychophysical law, one must always consider the effect of the series of stimuli being judged, as a frame of reference or context.

With this view as a guide Helson generated an extensive series of investigations to test a quantitative theory of the adaptation level (Helson, 1947; 1964a; 1964b) and his writings were featured in several major texts in the 1950's through the early 1980s, such as Bartley (1980), Dember and Warm (1979), Stevens (1951), Uttal, (1981); Woodworth & Schlosberg, 1954; Kling & Riggs, (1971), Osgood (1953), Graham (1965); Forgus & Meloamed (1976); Von Fieandt (1966), and Fiske and Maddi (1961). However, except for psychophysical monographs by Bolanowski & Gescheider (1991) and Gescheider (1997), current general perception texts such as Coren, Ward, & Enns, 2005; Sekuler & Blake (2002) and Schiffman (2000) place little emphasis on Helson's adaptation level concept.

Helson's Historical Role

1. The passage of time

In the 40 years since *Adaptation Level (1964a)* the role of context in Psychology has become so pervasive, we cannot begin to review it. Does this account for the decline in Helson's presence in the literature? The modern formulae are far advanced beyond his beginning equations, and there is no need to advocate for the study of context.

2. Rivalry with, S.S. Stevens,

Stevens was one of the main proponents of the *input-output* view of psychophysics. Posthumously, Steven's (1975, p 296) who did not appreciate Helson's demand to include the effects of context as part of experimental research, refers to a study by Garner (1953), as contrary fact to Helson's 1964 adaptation-level theory. Helson students, such as the first author, find this conclusion itself, contrary to Helson's own predictions as it ignores the issue of assimilation vs. contrast effects. Had he lived would Stevens have re-written this section? Stevens' students seem to support this passage without revision. Stevens also pointed out that adaptation level theory has nothing to do with the adaptation of a sense organ (Stevens, 1975 p297). To be sure, that claim is true, because adaptation level theory is concerned with cognitive effects on psychophysical judgments..

3. Empirical tests of Adaptation Level

Another possibility for the disappearance of Helson from the literature is that tests of the specific quantitative predictions from his adaptation level theory show only finite support. Baker et al., (1993) reported finding many research articles using Helson's concepts, with about 50% in support. A comparison with other historical figures is not easy. We generally accept the idea that the success of Weber, Fechner, Stevens and others falls far short of perfection. However, we accept the historical importance of their work in general, as there is historical merit to Helson's emphasis on the general principal of contextual bounds in psychophysics.

4. Current work on Contextual Influences

Weiler, E.M., Baker, M.A., Steger, J.A., & Sandman, D.E. (1995) found that previous work (Sandman, Weiler, & Pedersen, 1982) showing unexpected *consistency* in loudness sensory adaptation effects at *differing* intensities could be accounted for by Helson's theory of judgmental adaptation level. They also noted that Marks' (1994) concept of "recalibration" of judgment, derived independently of Helson's AL, represented a new explanation of the influence of context on loudness judgments. Prior work by Algol & Marks (1990) also provides additional modern ideas of context-bound psychophysics.

Along that line, Locked (2004) has recently provided an elegant argument against the simple **input-output** model and contributed a lucid account of the role of context in modern psychophysics. Professor Locked (2004) did not include Helson in his review, perhaps because his work is not sequential to Helson's but followed a different tradition. Nevertheless, for the historical record, we argue that Helson's 40 years of work merits recognition as a major historical influence to establish the contemporary *Zeitgeist* where context plays multiple roles in Psychological research.

References

Algom, D. & Marks, D.E. (1990). Range and regression, loudness scales, and loudness processing: toward a context-bound psychophysics. Journal of Experimental Psychology: Human Perception and Performance. 16, pp.706-727.

Allport, F.H. (1955). Theories of Perception and the Concept of Structure. Chapter 10, pp 242 ff. John Wiley & Sons, Inc. N.Y.

Baker, M.A., Avant, L.L., Steger, J., Weiler, E.M., Dember, W. & Warm, J. (1993). Helson Revisited: his current impact on psychophysics and psychology. Fechner Day 1993, proceedings of the ninth annual meeting of the International Society for Psychophysics. October 21-24, 1993. Pp.25-28.

Bartley S.H. (1980). Introduction to Perception. Harper & Row NY,NY

Bolanowski, S. J. Jr., & Gescheider, G. A. (1991). Ratio Scaling of Psychological Magnitude- In Honor of the memory of S.S. Stevens. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ

Coren, S., Ward, L.M., & Enns, J.T. (1999). Sensation and Perception. Fifth ed. Harcourt Brace College Publishers. pp.47-48.

Coren, S., Ward, L.M., & Enns, J.T. (2005). Sensation and Perception. Sixth ed. Harcourt Brace College Publishers. bibliography.

Corso, J.F. (1967). The Experimental Psychology of Sensory Behavior. Holt, Rinehart and Winston

Dember, W.N., & Warm, J.S. (1979). Psychology of Perception, 2nd ed. New York: Holt, Rinehart and Winston

Fiske, D.W. & Maddi, S.R. (1961). Functions of Varied Experience. The Dorsey Press, Homewood, Ill

Forgus, R.H. & Melamed, D.E. (1976) Perception: A Cognitive-Stage Approach (2^{nd} ed.) McGraw-Hill Book Company, NY, NY.

Gescheider, G.A. (1997). Psychophysics: The Fundamentals, 3d ed., Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ

Graham, C.H. Ed. (1965). Vision and Visual Perception. John Wiley & Sons, , NY,NY

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. American Journal of Psychology, 60, 1-29.

Helson, H. (1964a). Adaptation-Level Theory. Ch. 4. Harper & Row, NY, NY.

Helson, H. (1964b). Current trends and issues in adaptation-level theory. American Psychologist, 19, pp26-38.

Helson. H. (1959). "Adaptation-level theory" pp. 555-615. In S. Koch. (Ed.) Psychology: A Study of a Science. Vol I. Sensory, Perceptual, and Physiological Formulations. McGraw-Hill NY

Helson, H. (1964). Adaptation-level theory: an experimental and systematic approach to behavior. Harper & Row, publishers, NY,NY.
Kling, J.W. & Riggs, L. A. (2005). Woodworth & Schlosberg's Experimental Psychology. Third Ed. Holt, Rinehart and Winston, Inc. Pp.416.
Lockhead, G.R. (2004). Absolute judgments are relative: a reinterpretation of some psychophysical ideas. General Psychology. 8, p. 265-272.
Luce, R.D. & Suppes, P. (2002). Representational Measurement Theory. Chapter 1 (pp.1-41) in Stevens' Handbook of Experimental Psychology, Third Ed. Vol. 4: Methodology in Experimental Psychology, John Wiley & Sons, Inc.
Marks, D.E. (1994). "Recalibrating" the auditory system: the perception of loudness. Journal of Experimental Psychology: Human Perception and Performance, 20, 382-396.
Osgood, C.E. (1953). Method & Theory in Experimental Psychology, Oxford University Press, NY, NY.
Sandman, D.E., Weiler, E.M. & Pedersen, (1982). Perstimulatory adaptation by magnitude estimation: Effects of time of day;, level of continuous tone, and relative level of periodic increments. Journal of Auditory Research, 22, 65-69.
Schiffman, H.R. (2000). Sensation and Perception: An Integrated Approach. Fifth Ed. Pp.43. John Wiley & Sons, Inc. NY, NY.
Sekuler, R. & Blake, R. (1990). Perception. McGraw-Hill, Inc. pp. 458. NY, NY.
Steger, J.A., Weiler, E.M., Sandman, D.E., & Baker, M.A. (1999). Helson: from the gestalt to adaptation level theory. In Theme Session IV, Fechner Day 1999, Proceedings of the fifteenth annual meeting of the International Society for Psychophysics. October, 21-24, 1999. pp.68-71.
Stevens, S.S. (1975). Psychophysics: introduction to its perceptual, neural and social prospects. John Wiley & Sons, Inc. New York.
Stevens, S.S. (1958). Handbook of Experimental Psychology. John Wiley & Sons,. NY, NY
Uttal, W.R. (1981). A Taxonomy of Visual Processes. Lawrence Erlbaum Assoc. N.J. Pp.-84-85.
Von Fieandt, K. (1966). The World of Perception. The Dorsey Press, Homewood, Ill
Weiler, E.M., Baker, M.A., Steger, J.A., & Sandman, D.E. (1995). Using Helson's concept of adaptation level to reconsider recent research. Fechner Day 1995, Proceedings of the eleventh annual meeting of the International Society for Psychophysics. Oct 21-25, 1995. Pp.327-332.
Woodworth, R.S. & Schlosberg, H. (1954). Experimental Psychology. Second Ed. Holt, Rinehart and Winston Inc. pp. 231-233, 407, 448-449. NY, NY.

Perceived Exertion during Brief Isometric Abduction Contraction in the Plane of the Scapula.

Mark K. Timmons, Staci M. Stevens, & Danny M. Pincivero
University of Toledo, Toledo, OH, USA
mark.timmons@utoledo.edu

Abstract

Purpose: To explore the effects of increasing shoulder abduction angle on ratings of perceived exertion (RPE) during isometric abduction contractions. **Methods**: Twelve healthy subjects performed maximal voluntary (MVC) and sub-maximal (10%-90% MVC) isometric arm abduction contractions in the plane of the scapula, on a Biodex dynamometer. Abduction torque and electromyograms (EMG) of the deltoid and scapular rotator muscles were recorded during each contraction. Following each sub-maximal contraction the subject was asked to rate their perceptions of exertion during the contraction on a Borg CR-10 scale. Integrated EMG (IEMG) values were calculated during a 3 second window of each contraction. Sub-maximal IEMG was then normalized to the MVC IEMG. Separate repeated measures analyses of variance were conducted on the dependant variables RPE and normalized IEMG. **Results**: RPE and normalized IEMG were found to increase as arm abduction angle increased at contraction intensities of equal relative magnitude. **Conclusion**: RPE increased as a result of an increase in central drive as indicated by the increase in the IEMG.

Researchers in the areas of fatigue, rehabilitation and ergonomics frequently use the evaluation of a subject's perception of physical effort as a research tool. Subjects are asked to rate their exertion during long duration contractions or while doing repetitive motions in completing work related tasks. These techniques have been used in investigations involving the upper and lower body as well as whole body tasks. A decline in an individual's ability to sense physical effort may lead to an increase in the risk of injury or a reduction in physical performance.

Theories as to how we perceive physical effort are based on both central and peripheral control factors. Central sensation of effort is through a feed forward mechanism, where the output from the motor cortex is also sent to the sensory cortex (Cafarelli, 1982). Physiological processes associated with central sensation of effort are heart rate, respiratory rate and oxygen uptake. Peripheral sensation occurs through a feedback mechanism. Peripheral signals of effort arise from muscle, tendon and joint position receptors (Cafarelli, 1982). Initially sensation of exertion is controlled by peripheral mechanism and as exercise continues the sensation of exertion is modified by central mechanisms (Robertson, 1982).

Several methods of determining an individual's level of physical exertion have been developed over time. Each of these methods has strengths and weaknesses that lend each method to clinical and research activities (Borg, 1982). The Borg 10 category ratio scale (CR-10) was developed in order to place objective measure from subjective perceptions (Borg, 1982; Noble & Robertson, 1996). The Borg CR-10 scale allows for between subject comparisons of RPE. Ratings of perceived effort (RPE) have been shown to increase following both linear and non-linear trends with increasing muscle contraction intensities. In a study involving isometric knee extension contraction Pincivero et al (2000) found that the RPE on the CR-10 scale produced a linear increase with increasing contraction intensity.

Arm abduction requires synchronous upward rotations of the scapula and humerus. This change in the bony configuration of the shoulder girdle allows for the large range of motion of the shoulder and increases glenohumeral stability at the extremes of motion. With increasing arm abduction angle the muscles that rotate the scapula and abduct the humerus are shortened. Solomonow et al (1991) and Van Woesnel and Arwert (1993) have shown that during contractions of equal relative intensities, shorter muscle lengths will produce higher EMG values. The feed forward mechanisms that produce the higher EMG with decreased muscle length should also produce high RPE at equal relative intensity contraction at higher arm abduction angles. This investigation explored the effect of arm abduction angle on the perception of exertion and muscle activation during sub-maximal isometric shoulder abduction contractions.

Methods

Subjects

Subjects for this study consisted of 12 healthy adult volunteers (8 men, 4 women, age 23.8 ± 4.5 years, height 171.6 ± 10.1cm, mass 72.4 ± 15.9 kg) participated in this investigation. All subjects participated in recreational and/or occupational physical activity at least three times per week. No subjects were involved in either recreational or occupational activities that required excessive over head activity. No subjects had a history of traumatic shoulder injury, shoulder surgery or had a history of shoulder pain or pathology that required medical attention during the previous 2 years. All subjects gave their written informed consent to participate and the investigation was approved by the Human Subjects Research Review Board of The University of Toledo.

Procedure

Each subject participated in 6 testing sessions, separated by approximately 1 week. During each test session, subjects were assessed for shoulder abduction torque production during isometric contractions in the scapular plane and subjective ratings of exertion during sub-maximal isometric contractions. Each testing session began with a five minute warm-up on an upper body ergometer. Following the warm-up the subjects were seated in an upright position on the Biodex chair such that their right arm was aligned in the plane of the scapula, 45° anterior to the subject's frontal plane. Subjects then performed 5 isometric abduction maximal voluntary contractions (MVC). The subject was given a 2 minute rest period between contractions, during the rest period the subjects arm was removed from the test position. The subject then performed 9 sub-maximal contractions at 10-90% of their MVC, in predetermined random order. Each testing session differed by the arm abduction angle (15°, 30°, 45°, 60°, 75° and 90°); the order of testing positions and sub-maximal contractions were presented in a random order.

Isometric Maximal Voluntary Contractions

Maximum and sub-maximal isometric torque was measured with a Biodex Isokinetic Dynamometer (Biodex Medical System, Shirley). The subject was secured using pelvic and torso straps in order to minimize extraneous body movements. The axis of rotation of the Biodex was aligned with the axis of rotation of the subject's right shoulder. While seated, the subject's forearm was secured to the resistance adapter of the dynamometer proximal to the radial styloid. The subject was instructed to point their thumb towards the ceiling. Following 2-3 sub-maximal familiarization contractions, subjects then preformed 5 abduction MVC trials. Verbal encouragement was given during the MVC trials in order to assure the subjects maximal effort. Each MVC was held for 5 seconds. The mean torque was determined during a 3 second window of maximum torque development for each MVC trial. Mean abduction MVC was determined by averaging the 3 highest abduction MVC trials.

Isometric Sub-maximum Contractions

Immediately following the abduction MVC determination, the subject performed 9 sub-maximal isometric abduction contractions. Sub-maximal contractions were performed at 10, 20, 30, 40, 50, 60, 70, 80, and 90% of the mean MVC. A horizontal line was place on the Biodex monitor at the level of the target torque. Subjects were given verbal and visual instructions to assist with task of matching the target line. The Biodex monitor was masked such that the subject could not see the torque values. The subject was instructed to raise the cursor to match the horizontal line and maintain that position throughout the contraction. Sub-maximal contractions were held for 10 seconds.

Perceived Exertion

Perceived exertion was assessed immediately following each sub-maximal contraction, before the arm was removed from the test position. Perceptual high and low anchors were established during the MVC determination (Pincivero, et al, 1999). Prior to the initiation of each MVC with the arm positioned at the test abduction angle, the investigator displayed an enlarged copy of the modified Borg CR-10 scale to the subject while highlighting the "zero or nothing" descriptor, the subject was instructed to think of the feelings about their shoulder as "zero or nothing",. Immediately following each MVC the subject was again shown the modified Borg CR-10 scale while the investigator highlighted the maximum descriptor and instructed the subject to think of the feelings about their shoulder during the contraction as maximum. Immediately following each sub-maximal contraction with the subject's arm remaining in the test position, the subject was shown the Borg CR-10 scale and asked to rate their exertion level while thinking about their feelings about their shoulder during the contraction. Subjects were instructed that they could provide a number higher than 10 if they so desired.

Electromyography

During each contraction electromyograms (EMG) were obtained from the deltoid (DEL), upper trapezium (UT), lower trapezium (LT) and serratus anterior (SA) muscles. Electrodes were placed as to Basmajian (Basmajian, 1983), DEL midway between acromium process and the deltoid tuberosity, UT one third the distance between C7 and the lateral acromium process, LT half the distance between T7 and the inferior angle of the scapula and SA at a point along the mid-axillary line over the 6^{th} rib. EMG signals were collected at 2000 Hz, bandpass filtered (20-500 Hz) and fullwave rectified. A 3 second analysis window was chosen during the point of maximal torque for the MVC's and during the point at which the subject was producing the target torque. The integrated EMG (IEMG) value was calculated for each analysis window. IEMG from the sub-maximal contractions were normalized (NORM IEMG) to the MVC.

Statistical Procedures

Means and standard deviations were calculated for contraction intensity and arm angle. A one sample t-test was performed testing that the mean perceived exertion for of the each contraction intensities differs from the target contraction intensity. A two factor (contraction intensity x abduction angle) repeated measures analysis of variance (ANOVA) was performed on the independent variable RPE. A three factor (muscle x contraction intensity x abduction angle) repeated measures ANOVA was performed on the dependant variable IEMG. Statistical significance was determined at $p < 0.05$.

Results

The MVC torque decreased significantly ($F_{(5, 40)} = 7.44$, $p<0.001$) as AA increased presented in table 1. Ratings of perceived exertion increased as CI increased, across all AA, were as follows: 1.0, 1.8, 2.7, 3.8, 5.0, 6.1, 7.0, 8.1, and 9.4 for CI ranging from 10-90% MVC, respectively. RPE for each AA and CI are presented in table 2. Repeated measures ANOVA performed on the dependant variable RPE demonstrated significant main effects for CI ($F_{8,88} = 543.89$, $p<0.001$), and AA ($F_{5,55} = 2.78$, $p = 0.03$). The CI by AA interaction was not statistically significant ($F_{40,440} = 1.12$, $p = 0.285$). Pair-wise comparisons showed significant differences amongst all CI levels. RPE by AA are presented in table 3. Pair-wise comparisons between AA levels showed significant differences between the 15° (mean 4.6 ± 0.2) and the 75° (mean 5.3 ± 0.2, t=3.13, p=0.010) and 90° (mean 5.2 ± 0.2, t=2.26, p=0.045) AA positions, the 30° AA (mean 4.8 ± 0.2, t=2.35, p=0.04) position was significantly different from the 75°AA position.

Table 1. MVC at each arm abduction angle. * Statistically different from lower arm abduction angles p < 0.05.

Abduction Angle (degree)	Moment (Nm)
15	58.25 ± 6.14
30	57.97 ± 6.71
45	53.08 ± 5.38*
60	50.08 ± 5.38*
75	47. 83 ± 5.52*
90	47.51 ± 6.32

Table 2, Ratings of perceived exertion by abduction angle (degree) and contraction intensity (%MVC). * Statistically different form CR-10 level (1-9), p <0.05

	Contraction Intensity								
Arm Angle	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**
15	0.5±0.5*	1.3±0.8*	1.9±0.5*	3.0±1.1*	4.9±1.2	5.9±1.1	6.8±1.3	8.3±0.9	9.1±0.9
30	1.1±0.8	1.6±0.5*	2.5±1.1	3.4±0.9*	4.9±1.1	6.1±1.6	6.8±1.2	7.8±1.7	8.9±1.1
45	1.1±0.8	2.2±0.7	2.7±0.9	3.8±1.1	4.8±0.8	6.3±1.3	7.2±1.3	7.5±2.4	9.4±1.4
60	1.1±0.3	1.7±0.5*	2.6±0.8	4.3±0.7	4.8±0.8	6.1±1.0	7.1±1.1	8.2±1.4	9.3±1.1
75	1.1±0.7	2.4±1.2	3.1±0.9	3.9±0.7	5.3±1.1	6.0±1.6	7.4±1.2	8.6±1.2	10±1.3*
90	1.4±0.7	1.9±0.8	3.1±1.1	4.3±1.3	5.0±0.9	6.2±1.3	6.8±1.2	8.3±0.8	9.7±1.0*

Table 3, Mean RPE for each AA, * statistically different from 75 and 90 degree levels, ** statistically different from 75 degree level.

Abduction Angle (degree)	Mean RPE
15	4.6 ± 0.15*
30	4.7 ± 0.23**
45	5.0 ± 0.20
60	5.0 ± 0.16
75	5.3 ± 0.21
90	5.19 ± 0.18

Normalized IEMG increased as AA increased, Figure 1. Repeated measures ANOVA of the normalized IEMG dependant variable showed significant main effect for both AA ($F_{5, 55} =13.67$, $p<0.001$) and CI ($F_{8, 88} = 400.81$, $p<0.001$) and significant interaction for AA by CI ($F_{40, 440} = 2.41$, $p<0.001$) and muscle by CI ($F_{24, 264} = 1.67$, $p=0.029$). The lower AA levels (15, 30, and 45 deg) were each significantly different for the higher AA levels (60, 75, and 90 deg). The muscle main effect and the three way interaction effect were found to be statistically insignificant.

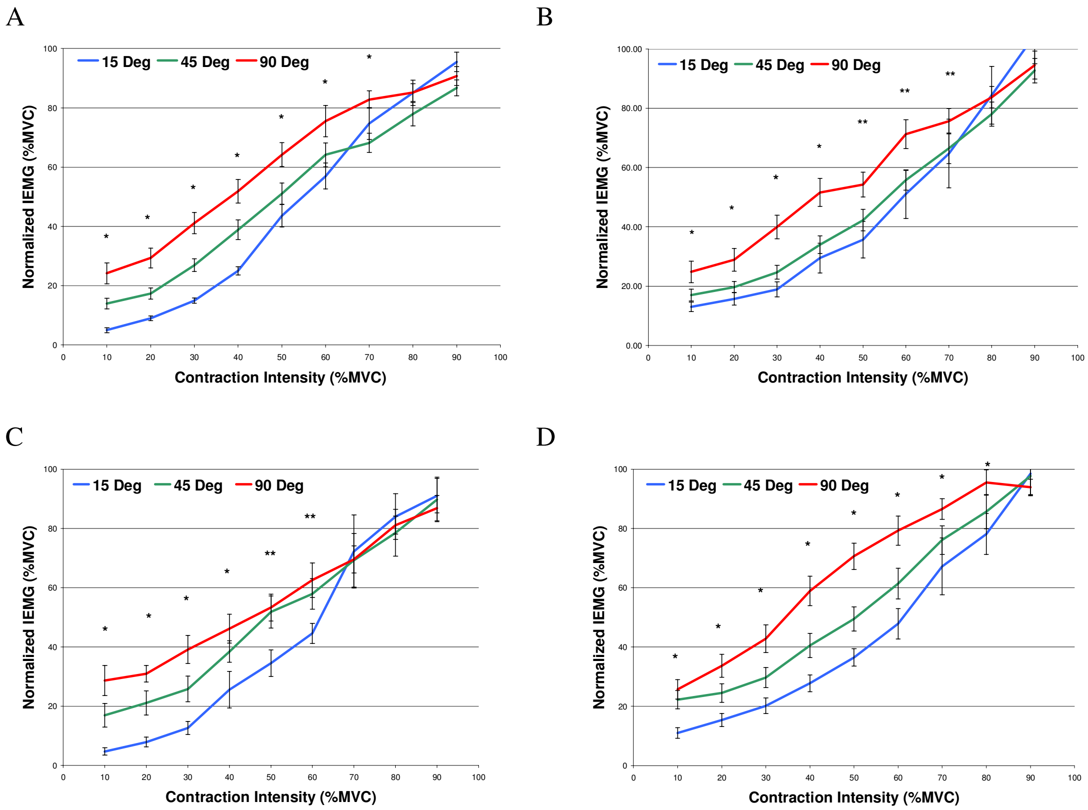


Figure 1, Normalized IEMG as a function of contraction intensity. A, Deltoid B, Upper Trapezium C, Lower trapezium D, Serratus Anterior.

Discussion and Conclusions

The decrease in muscle length produced by increasing arm abduction angle increased normalized IEMG values at the lower relative contraction intensities and RPE. At lower CI increases in the magnitude of the EMG signal are primarily due to the increase in number of motor units recruited (Woods & Bigland-Ritchie, 1983). We must assume that at near maximum contraction intensities nearly all motor units within a given muscle would be activated. Increases in the normalized IEMG at higher AA would represent an increase in the total number of motor units recruited than lower abduction angles. We also found that at higher AA lower abduction torque was produced, so more motor units were creating less abduction torque. The increase in IEMG at higher AA would be produced by stronger central control input to the muscle. Peripheral input from the muscles investigated would be minimized due to the shortening of these muscles as AA increased. Lower abduction torques would suggest lower joint forces and less tension within the muscle and tendon, lower tension should reduce the amount of peripheral input at higher AA. The muscles investigated in this study would all shorten as AA increases it is possible that the opposite relationships between IEMG and RPE would be seen if muscles that lengthen during shoulder abduction were studied.

References

Basmajian, J. (1983). *Biofeedback Principles and Practices for Clinicians* (2 Ed.). Baltimore, MD: Williams and Wilkins.

Borg, G. (1982). Psychophysical bases of perceived exertion. *Medicine and science in sport and exercise,* *14*(5), 377-381.

Cafarelli, E. (1982). Peripheral contributions to perception of effort. *Medicine and science in sport and exercise, 14*(5), 362-369.

Noble, B., & Robertson, R. (1996). *Perceived Exertion*. Champaign, IL: Human Kinetics Publishers.

Pincivero, D., Coelho, A., & Erikson, W. (2000). Perceived exertion during isometric quadriceps contraction. *Journal of Sports Medicine and Physical Fitness, 40*, 319-326.

Pincivero, D., Lephart, S., Moyna, N., Karunakara, R., & Robertson, R. (1999). Neuromuscular activation and RPE in the quadriceps at low and high isometric intensities. *Electromyography and Clinical Neurophysiology, 39*, 43-48.

Robertson, R. (1982). Central signals of perceived exertion during dynamic exercise. *Medicine and Science in Sport and Exercise, 14*(5), 390-396.

Solomonow, M., Baratta, R., & D'Ambrosia, R. (1991). EMG-force relations of single muscle acting across a joint: dependence on joint angle. *Journal of Electromyography and Kinesiology, 1*(1), 58-67.

Van Woensel, W., & Arwert, H. (1993). Effects of external load and abduction angle on EMG level of shoulder muscles during isometric action. *Electromyography and Clinical Neurophysiology, 33*, 185-191.

Woods, J., & Bigland-Ritchie, B. (1983). Linear and non-linear surface EMG/force relationships in human muscles. *American Journal of Physical Medicine, 62*(6), 287-299.

ON INTERNAL TIME MEASURES AND PROPERTIES OF THE KLEPSYDRAIC CLOCK

Jiří Wackermann

Dept. of Empirical and Analytical Psychophysics
Institute for Frontier Areas of Psychology, Freiburg i. Br., Germany

jw@igpp.de

Abstract

In the present paper we examine properties of internal time representation based on lossless integration of time-varying flows. We find that for flow functions of exponential type the lossless integration model is equivalent with our 'dual klepsydra model' proposed earlier. Further we examine properties of time measures generated by successive reproduction of a unit standard, and their relations to the reproduction functions. Finally, we demonstrate some special features of the 'klepsydraic clock', supporting its presumably outstanding role as a model of internal time representation and measurement.

The problem of internal representation of temporal relations between external events, and its role in organismic behaviour and subjective experience, is essentially a metricisation problem. By this virtue, the problem presents a challenge not only to neurobiology or neuropsychology, but particularly to mathematical psychology and psychophysics.

Time representation: clocks as flow integrators

Time is the dimension of experience relative to which events can be uniquely ordered. The temporal order is given by an irreflexive, asymmetrical, and transitive *precedence* relation '$\prec$'. Two events A and B, for which neither $A \prec B$ nor $B \prec A$ is true, are *simultaneous*, $A \sim B$; simultaneity is a reflexive, symmetrical, and transitive, *i. e.*, equivalence relation. It is further postulated that the relation '$\prec$' is *universal*, *i. e.*, for any two events A and B one (and only one) of propositions $A \prec B$, $A \sim B$, $B \prec A$ holds (*cf.* Le Poidevin, 1993).

Clocks are physical processes, generating series of events $\{C_n\}$ such that $C_n \prec C_{n+1}$ for all $n \in \mathbb{Z}$. A certain class of congruent[1] clocks, $\mathbf{T}$, is chosen, and their readings identified with universal time data $t_n = t_0 + n\,P$, where P is a period constant for a given clock. Constancy of P is no empirical fact—it is a convention (Poincaré, 2003; Reichenbach, 1957), but not entirely an arbitrary one. Its choice is dictated by economical description (Mach, 1960) of elementary physical processes, particularly of those underlying periodic clocks themselves, *e. g.* pendulum isochrony. In Huygens' clock, the swinging pendulum generates discrete events ('ticks') at a constant frequency; the number of ticks accumulated during a certain time interval $[t_1; t_2]$ serves as a measure of the duration $t_2 - t_1$.

The *internal clock model* (i. c. m.) (Treisman, 1963; Church, 1984) is, in principle, an huygensian clock, operating with a 'pacemaker' and 'pulse counter'. Obviously, a clock operating at a constant pacemaker frequency will attain equal pulse counts within equal times after counter reset, so such a clock would exactly reproduce time intervals, regardless of its internal pacemaker frequency. This is at odds with systematic deviations of reproduced durations from the chronometrically correct response, observed in experimental data (Fig. 1). Thus at least one of the two assumptions must be abandoned, (i) lossless accumulation, and/or (ii) constant pacemaker frequency.

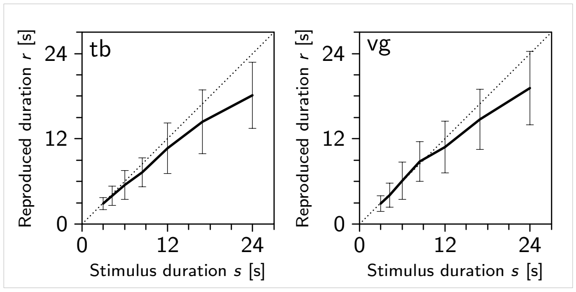


Fig. 1. — Individual time reproduction curves from two different subjects. The data were obtained in multiple experimental sessions (six with subject **tb**, five with subject **vg**), each session consisting of 5 repetitions for each of 7 different stimulus durations s (Wackermann, Späti, & Ehm, 2005). Shown are arithmetic means ± 1 s. d. — Note the relative shortening of reproduced times r with increasing stimulus duration s; the dotted line indicates the 'chronometrically correct' response $r = s$.

If the pacemaker period is much shorter than typical durations measured, we may assume a continuous 'flow' density, f, which is the number of pulses per time unit, and replace summation by integration. Generally, we postulate an inflow/outflow system (i. o. s.) described by an ordinary differential equation

$$\mathrm{d}y = (f(t) - g(y))\,\mathrm{d}t\,,$$

where f represents the inflow as a function of time, and g is a function relating the outflow to the momentary state y. In our '*dual klepsydra model*' (d. k. m.) (Wackermann, Ehm, & Späti, 2003) of time reproduction, we suggested a system consisting of two i. o. s., with $f(t) = \mathrm{const}$ and a linear 'leakage' term $g(y) = \kappa\, y$. The d. k. m. fits experimental data with good accuracy and has certain appealing analytical properties (Wackermann, 2005). Here we study properties of time representation based on lossless ($g \equiv 0$) integration of time-varying flows.

Duration reproduction with time-varying flows

Since f is a positive function of time $t \geq 0$, its integral

$$F(t) \equiv \int_0^t f(\tau)\,\mathrm{d}\tau$$

is a monotonically increasing, and thus invertible, function. For later use we define two new operations on $\mathbb{R}^2_{0+}$, 'F-addition' and 'F-multiplication', as follows:[2]

$$x \oplus_F y \equiv F^{-1}(F(x) + F(y))\,, \quad k \otimes_F x \equiv F^{-1}(k\,F(x))\,.$$

Further, let $s > 0$ be the presented duration, $w \geq 0$, the pause between the end of presentation and the beginning of reproduction, and r, the reproduced duration. During the presentation phase, $0 \leq t \leq s$, the flow f is integrated in an accumulator. The state of the accumulator is then stored in a 'memory register', the accumulator is reset to zero, and reused to integrate the flow f during the reproduction phase, $s + w \leq t$. The reproduction ends when the accumulator state reaches the state attained during the presentation phase, at time $t = s + w + r$. The reproduction condition is thus

$$F(s + w + r) - F(s + w) = F(s) - F(0)\,, \tag{1}$$

where, of course, $F(0) = 0$. Rearranging terms and applying the inverse F^{-1} to both sides of the equation, we obtain for the reproduced time r, expressed as a function of s and w,

$$r(s, w) = s \oplus_F (s + w) - (s + w)\,. \tag{2}$$

Table 1 gives an overview of the results for three most important types of flow density functions. (The constant flow, #1, is a special case of the next two types, #2 and #3, but is shown separately for

Table 1. Reproduction functions for three special flow function types.

Flow function type	Density $f(t)$	Integral $F(t)$	Reproduction function $r(s,w)$
#1 Constant	c	$c\,t$	s
#2 Power function*	$c\,(\frac{t}{a})^{\beta-1}$	$\frac{c\,a}{\beta}(\frac{t}{a})^{\beta}$	$s\,\left(\left(1+(1+\frac{w}{s})^{\beta}\right)^{1/\beta}-(1+\frac{w}{s})\right)$
#3 Exponential†	$c\,\mathrm{e}^{\lambda t}$	$c\,\lambda^{-1}\left(\mathrm{e}^{\lambda t}-1\right)$	$\lambda^{-1}\log\left(1+\left(1-\mathrm{e}^{-\lambda s}\right)\mathrm{e}^{-\lambda w}\right)$

* $c, a > 0$, $\beta > 1$, reduces to case #1 for $\beta \to 1$ † $c, \lambda > 0$, reduces to case #1 for $\lambda \to 0$

the reader's convenience.) Of particular interest are exponential flows, yielding a reproduction function identical to the 'klepsydra reproduction function' for the linear d. k. m.,

$$\mathrm{krf}(s,w) = \kappa^{-1}\log\left(1+\eta\,(1-\mathrm{e}^{-\kappa\,s})\,\mathrm{e}^{-\kappa\,w}\right)\,,$$

with $\kappa = \lambda$ and $\eta = 1$. The two models, the d. k. m. based on lossy accumulators with constant inflow, and the i. c. m. based on lossless accumulation of exponential inflow, are *observationally equivalent*, *i. e.*, not distinguishable in terms of empirical data.

Time accounting: iterated reproduction of unit

First we briefly review the concepts of fundamental measurement (Campbell, 1957). Measurement of a property X is defined by a structure $\mathfrak{M}_X \equiv \{\mathbf{U}, \varphi, \sqcup, u\}$, where $\mathbf{U}$ is a universe of physical objects, $\varphi : \mathbf{U} \to \mathbb{R}_{0+}$ is a function assigning numerical values to the objects, $\sqcup : \mathbf{U}\times\mathbf{U} \to \mathbf{U}$ is an 'aggregation operation'[3] such that $\varphi(a \sqcup b) = \varphi(a) + \varphi(b)$ for $\forall a, b \in \mathbf{U}$ (additivity axiom), and $u \in \mathbf{U}$ is a chosen 'unit' standard, $\varphi(u) = 1$.

The above is just the minimal logical structure of measurement; physical realisation of measurement principally requires the *reproducibility* of the unit. For example, in case of spatial measurement, 'u' may be a rigid rod used as the unit standard, and '$\sqcup$' is a linear concatenation of lengths in a given direction. Let '$\sqsubset$' denote the relation 'covered by'. Given a length x, we have to determine the minimal number, say n, of units needed to cover x,

$$\underbrace{u \sqcup \ldots \sqcup u}_{(n-1)\times} \sqsubset x \sqsubseteq \underbrace{u \sqcup \ldots \sqcup u}_{n\times}\,,$$

which implies $n - 1 < \varphi(x) \leq n$. These n unit length segments may be (i) multiple copies of the unit manipulated *simultaneously*, or (ii) marked positions of a single unit standard which was *successively translated* in the given direction.[4] In case of measurement of *temporal durations*, the formal structure is the same, but only option (ii) is available for physical realisation: the measured time interval is to be covered by successively reproduced clock periods.[5]

Klepsydraic clock and its properties

Let $u > 0$ be a given 'unit' duration, and r a reproduction function. We define a sequence $\{{}^{\mathrm{o}}r_n\}_{n\in\mathbb{N}}$ of iterative reproductions of the *original unit*,

$${}^{\mathrm{o}}r_{n+1} = r\left(u, {}^{\mathrm{o}}T_n\right)\,, \quad \text{where } {}^{\mathrm{o}}T_0 \equiv 0 \text{ and } {}^{\mathrm{o}}T_n \equiv \sum_{\nu=1}^{n} {}^{\mathrm{o}}r_\nu\,. \qquad (3)$$

Combining eqs. (2) and (3), it is easy to show that

$${}^{\mathrm{o}}T_n + u = (n+1) \otimes_F u\,,$$

so the coverage times ${}^{\mathrm{o}}T_n$ can be considered as 'generalised multiplication' of the unit.

Table 2. Times covered by iterative reproductions of unit u.

Flow function type	${}^{\mathrm{o}}T_n$	${}^{\mathrm{i}}T_n$	Note
#1 Constant	$n\,u$	$n\,u$	
#2 Power function	$n^{1/\beta}\,u$	$\sum_{\nu=1}^{\nu=n} \gamma^{\nu}\,u$	$\gamma \equiv 2^{1/\beta} - 1$
#3 Exponential	$\lambda^{-1} \log(1 + n\,\epsilon)$	$\lambda^{-1} \log(1 + n\,\epsilon)$	$\epsilon \equiv 1 - \mathrm{e}^{-\lambda u}$

Further we define a sequence $\{{}^{\mathrm{i}}r_n\}_{n\in\mathbb{N}}$ of iterative reproductions of the *unit's image*,

$$ {}^{\mathrm{i}}r_{n+1} = r\left({}^{\mathrm{i}}r_n, 0\right) \qquad \text{starting with } {}^{\mathrm{i}}r_0 \equiv u\,. \tag{4}$$

Elements of seq. (4) can be expressed as iterations of the partial function $r_0 : s \mapsto r(s, 0)$, so that ${}^{\mathrm{i}}r_n = r_0^n(u)$, and summed up to ${}^{\mathrm{i}}T_n$. The resulting coverage times are shown in Table 2.

We can see that ${}^{\mathrm{o}}T_n = {}^{\mathrm{i}}T_n$ for all $n \in \mathbb{N}$ for exponential flows (and constant flows as their special case), but generally not for other flow types, *e. g.*, the power function with $\beta \neq 1$. Consequently, successive reproductions of the unit according to #3 (exponential flow), can be realised by a (bio)physical mechanism 'forgetting' the original unit, and the sequence ${}^{\mathrm{i}}T_n$ is identical with ${}^{\mathrm{o}}T_n$, thus providing an unambiguous time measure. (In the rest of this section we are discussing only the exponential flow case and its equivalent, klepsydraic clock, so we drop the superscripts 'o' and 'i' as unnecessary.)

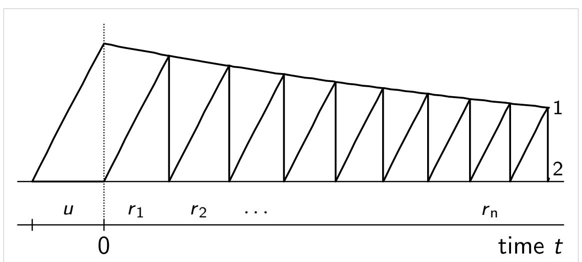


Fig. 2. — Implementation of a klepsydraic clock by a relaxation oscillator. Klepsydra 1 is initialised to a state corresponding to the unit u and continuously leaking afterwards. Klepsydra 2, periodically filled and reset to zero, generates a series of unequal clock periods r_n.

Because the i. c. m. with exponential flows and the d. k. m. are equivalent, all that has been said above applies to the d. k. m. as well. A 'klepsydraic clock' derived from the d. k. m. of unit reproductions is nothing else but a relaxation oscillator (Fig. 2), so that the average load of the i. o. s. involved decreases with time. This is compatible with a (neuro)physiologically realistic implementation of the model (*cf.* Wackermann, 2005, pp. 203–205). In contrast, the equivalent lossless integrator model requires exponentially increasing pulse flow density—a serious problem for biophysical implementation because organismic resources are naturally limited. This is a strong argument in favour of the lossy integration principle, and for the dual klepsydra model of duration reproduction.

Another remarkable property of the klepsydraic clock is its *internal consistency*. It is easy to prove that for any three integers $l \leq m \leq n$ the following identity[6] holds:

$$ \mathrm{krf}(T_m - T_l, T_n - T_m) = T_{n+m-l} - T_n\,. \tag{5}$$

This means that the 'klepsydraic measure' of a time interval, *i. e.*, number of 'klepsydraic periods' required to cover the interval, is *invariant* w. r. t. the klepsydraic reproduction. Transposed to length measurements, eq. (5) would express the invariance of lengths w. r. t. translation of the entire measurement setup (*i. e.*, the measured object along with the measurement ruler), which we normally take for granted in space *and* time. Our results show that this invariance can be preserved even in (bio/psycho)physical systems with a 'lossy' internal time representation, generating time measure which is incongruent with the universal time $\mathbf{T}$.

Concluding remarks

We have examined properties of lossless flow integrators with time-varying flows, as revealed by reproduction (duration $\mapsto$ duration), and production (numeric value $\mapsto$ duration) functions. We have shown that (i) reproductions with internal clocks based on exponential flows are equivalent to the formerly proposed 'dual-klepsydra model', and that (ii) productions obtained with the klepsydraic clock are invariant w. r. t. translation in time, *i. e.*, delayed reproduction. These results establish a connection from time reproduction models to models of unit-based time measurement, *i. e.*, to experimental 'time production' and 'estimation' tasks.

We have arrived at a point where we have models equivalent in terms of observational data, but different as to their analytical properties and physical realisation. Here we should note that in model competition, empirical criteria (*i. e.*, goodness-of-fit to experimental data) and meta-empirical criteria (*e. g.*, analytical simplicity and self-consistency) should keep parity. Studying problems of time metricisation, we are touching fundamental problems of the role of time in description of Nature, and the nature of our experience of time. Accordingly, *a priori* arguments[7] and axiomatic approches[8] shall play more important role than usually admitted by empirical researchers.

Acknowledgements

The author wishes to thank Robert C. Bishop and Werner Ehm for their critical comments and helpful suggestions on earlier drafts of the present paper.

Notes

[1] We call two clocks 'congruent' if their readings can be linearly transformed to each other; this is obviously an equivalence relation.

[2] F-addition is a commutative and associative operation, with a neutral element 0. To have F-addition defined on the entire domain $\mathbb{R}^2_{0+}$, F must be unbounded from above, *i. e.*, $\lim_{t\to\infty} F(t) = \infty$.

[3] The aggregation operation consists in some physical manipulation of the objects of measurement, and is fundamental for the definition of the property to be measured (Bridgman, 1927); whereas the choice of the unit standard is quite arbitrary, and the function φ is induced by those two choices.

[4] These unit segments are usually marked on a solid ruler, which is apposited to the measured length; thereby the reproducibility problem is not solved, it is only imposed on the manufacturer of the ruler.

[5] The logical rigidity of standard conditions for reproducibility of consecutive clock periods thus replaces the material rigidity of standard measurement rods (Janich, 1980, p. 191; *cf.* also Milne, 1940, p. 138).

[6] Note that formally $-T_{-1} = u$, as follows from Table 2, entry #3. The identity of reproductions of the original unit and the unit's image can thus be rewritten as $\mathrm{krf}(-T_{-1}, T_n) = T_{n+1} - T_n$, which is a special case of eq. (5) for $l = -1$ and $m = 0$.

[7] "[O]nly a parody portrays the scientist as merely accumulating facts and making generalizations on the basis of them. The 'raw' data, and the formulae which putatively describe them, require interpretation. Abstract ideas and a priori principles are never far away. [. . .] [There is] no reason to suppose either that the scientist will be unaffected by a priori principles or that the philosopher will be indifferent to empirical findings." (Le Poidevin, 1993, p. 162).

[8] "There are [. . .] two separate ways of attacking the structure of the universe: the one to make use of every available piece of empirical knowledge known to be valid on the small scale; the other to begin with the situation actually presented to us by the totality of things without supposing ourselves to know anything, to start with, of the facts used in the first method. In the second method we attempt a complete reconstruction of physics from the bottom upward, on an axiomatic basis." (Milne, 1940, p. 133).

References

Bridgman, P. W. (1927). *The Logic of Modern Physics.* New York: MacMillan.

Campbell, N. R. (1957). *Foundations of Science. The Philosophy of Theory and Experiment.* New York: Dover. [Original work: *Physics: The Elements*, 1920.]

Church, R. M. (1984). Properties of the internal clock, *Annals of New York Academy of Sciences*, **423**, 566–582.

Huygens, C. (1673). *Horologium oscillatorium, sive de motu pendulorum ad horologia aptato demonstrationes geometricae.* Paris: Muguet.

Janich, P. (1980). *Die Protophysik der Zeit. Konstruktive Begründung und Geschichte der Zeitmessung.* Frankfurt: Suhrkamp.

Le Poidevin, R. (1993). Relationism and temporal topology: physics or metaphysics? In R. Le Poidevin & M. MacBeath (Eds.), *The Philosophy of Time*, Oxford: Oxford University Press, pp. 149–167.

Mach, E. (1960). *The Science of Mechanics. A Critical and Historical Account of Its Development.* La Salle: Open Court, 1960. [Original work: *Die Mechanik, in ihrer Entwicklung historisch-kritisch dargestellt*, 1883.]

Milne, E. A. (1940). Cosmological theories. *Astrophysical Journal*, **91**, 129–158.

Poincaré, H. (2003). Simultaneity and the measure of time. In L. M. Dolling, A. F. Gianelli, G. N. Statile (Eds.), *The Tests of Time. Readings in the Development of Physical Theory*, Princeton: Princeton University Press, pp. 290–298. [Original work: 'La mesure du temps', 1898.]

Reichenbach, H. (1957). *The Philosophy of Space and Time.* New York: Dover, 1957. [Original work: *Philosophie der Raum-Zeit-Lehre*, 1927.]

Treisman, M. (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock". *Psychological Monographs*, **77**, whole No. 576.

Wackermann, J., Ehm, W., & Späti, J. (2003). The 'klepsydra model' of internal time representation. In B. Berglund & E. Borg (Eds.), *Fechner Day 2003*, Stockholm: International Society for Psychophysics, pp. 331–336.

Wackermann, J. (2005). Experience of time passage: Phenomenology, psychophysics, and biophysical modelling. In R. Buccheri, M. Saniga, & A. Elitzur (Eds.), *Endophysics, Time, Quantum, and the Subjective*, Singapore: World Scientific, pp. 189–208.

Wackermann, J., Späti, J., & Ehm, W. (2005). Individual response characteristics in time reproduction and time production tasks, *this volume*.

INDIVIDUAL RESPONSE CHARACTERISTICS IN TIME REPRODUCTION AND TIME PRODUCTION TASKS

Jiří Wackermann*, Jakub Späti*, Werner Ehm**

* *Dept. of Empirical and Analytical Psychophysics*

** *Dept. of Theory and Data Analysis*

Institute for Frontier Areas of Psychology, Freiburg i. Br., Germany

jw@igpp.de

Abstract

Results of a comparative study of time reproduction and time production are reported. Twelve subjects participated in a total of 59 experimental sessions. Reproduced or requested durations varied in the range from 3 to 24 seconds. Individual reproduction curves showed in most cases negative curvature, indicating progressive shortening of responses with increasing stimulus durations, while individual production curves were approximately linear functions of requested durations. Variances of reproduced or produced durations could be approximated by power functions of the stimulus magnitude; the characteristic exponents for reproduction variances were mostly lower than those for production variances. No significant correlations were found between parameters of individual reproduction and production curves. These results do not support the model of duration reproduction as a composite process, estimation → (internal datum) → production, but rather corroborate the concept of reproduction as an elementary process.

In the history of empirical research into 'time perception', several different experimental methods have developed (Woodrow, 1951; Bindra & Waksberg, 1956; Zakay, 1990). The major four methods can be divided into two groups, (i) *homogeneous*, operating exclusively with durations of sensorily perceivable events: reproduction and comparison methods; (ii) *heterogeneous*, translating the perceived duration into a symbolic/numeric representation, or *vice versa*: estimation and production methods.

It is often assumed that these methods measure the same quantity, 'subjective duration', only by different means—similarly as a physicist measures, say, intensity of electrical current by means of its various observable manifestations, *e. g.*, magnetic, thermic, or chemical effects. This assumption, however, is not sufficiently supported; it is well possible that different experimental methods may *activate different processes* employed by the subjects to cope with experimental tasks. Given the plenitude of paradigms and approaches to study in time experience and timing behaviour (*cf.* Grondin, 2001), there is no clear notion of what 'subjective duration' actually is, or how it can be measured.

Contemporary research shows a strong bias towards heterogeneous methods, particularly, to the *production* method. Results obtained with these methods are seemingly easy to interpret in terms of pacemaker frequency, assuming the 'internal clock model' (i. c. m.; see Church, 1984). However, the i. c. m. with a constant pacemaker frequency does not explain the progressive shortening of *reproduced* times observed at longer stimulus durations (*e. g.*, Vierordt, 1868; Eisler, 2003). Models with time-varying pacemaker frequency can, in principle, account for these phenomena in time reproduction data, but such models should also consistently predict production or estimation data—a challenge for future modelling efforts (*cf.* Wackermann, 2005).

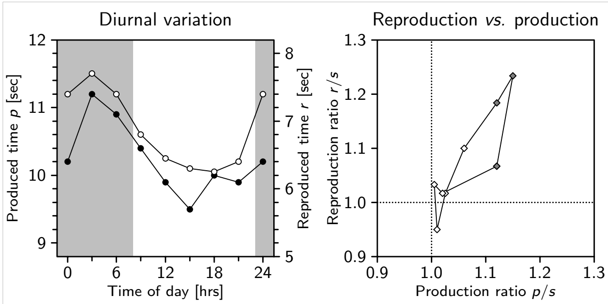


Fig. 1. Time production and reproduction data redrawn and reconstructed from Pöppel and Giedke (1970, p. 193, Fig. 7). —Left: Diurnal variation of produced times p (open circles, requested duration s = 10 s), and reproduced times r (filled circles, standard duration s = 6 s). Shaded areas indicate subject's sleep times. —Right: Covariation of production and reproduction data, plotted as ratios p/s and r/s, in the subject's diurnal cycle. Grey-filled markers indicate nocturnal measurements.

To address this problem, data collected repeatedly with the same subjects, using different methods in parallel, are needed. However, such studies are rather rare, and their interpretation is often unclear. For example, Pöppel and Giedke (1970), in a study of diurnal variations in time perception, reported time production and time reproduction data obtained by parallel measurements in one subject; the responses in those two tasks strongly correlated (Fig. 1). These results obviously require a more sophisticated interpretation than merely slow diurnal variations of 'locally constant' pacemaker frequency. Regrettably, the data were acquired with only one 'standard' duration for each task, hence provide no information about the *forms* of response functions and their variations.

The purpose of the reported study[1] was to systematically examine relations between responses in time reproduction and production tasks, and their correlations with the participants' diurnal sleep/waking cycle.[2] We are thus dealing with two sources of variance: (i) inter-individual (habitual) differences, and (ii) intra-individual, state-dependent variations. Only relations observable on level (i) are reported in the present communication.

Materials and methods

Apparatus

The tasks were presented on a portable computer, equiped with a 15-inch LCD screen and a pointing device (mouse) attached. Visual stimuli and instruction texts were displayed at the centre of the screen, at graphics display resolution 800 × 600 pixels, and watched from an average distance ≈45 cm. The visual stimulus indicating presented or (re)produced duration ('carrier') was a white circle of 40 pixels diametre (≈12 arc min), shown on a black background. Instruction texts between trials/tasks were displayed with Lucida Sans font of 40 points height.

Experimental procedure

Twelve paid volunteers participated in the study (7 female, 5 male, age range 19–44 years, mean age 30 years). The purpose of the study was explained to the participants, and their written consent obtained. Each subject participated in several experimental sessions during the day, repeated in a period of approximately two-hours; the total number of sessions depended on the subject's individual schedule of day-time activities. In each session, three tasks were given to the subjects: time reproduction task (t. r. t.), and time production task (t. p. t.), and a simple reaction time (r. t.) task.

In the *reproduction task*, the subjects had to reproduce a duration for which the carrier stimulus was shown on the screen. The stimulus was re-displayed after a constant inter-stimulus interval (i. s. i. = 1.25 s); the subject's task was to press the button after a time period subjectively perceived as equal to the presented duration. The stimulus disappeared at the button press, so the instruction was to reproduce the 'show-up time' of the carrier stimulus on the screen. Seven different stimulus durations were used, 3, 4.2, 6, 8.4, 12, 16.8, and 24 s (geometrical sequence with modulus $\sqrt{2}$), with 5 repetitions for each duration, presented in random order. A single t. r. t. run thus consisted of 7 × 5 = 35 trials.

In the *production task*, the subjects had to produce a specified duration of a given number of seconds. The numeric stimulus was displayed for 2 s. After the i. s. i. = 1.25 s, the carrier stimulus was displayed; the subject's task was to press the button when the requested duration, relative to the stimulus occurrence, had elapsed. Four different durations were used (a subset of durations used for the t. r. t., which could be expressed by integer numbers: 3, 6, 12, and 24 s), with 5 repetitions for each duration, presented in random order. A single t. p. t. run thus consisted of $4 \times 5 = 20$ trials.

The two time perception tasks thus differed only in the manner how temporal durations were presented to the subjects, as a visually perceived event (t. r. t.), or as a symbolic datum (t. p. t.). All subjects had normal or corrected-to-normal vision, all were right-handed, and operated the response button with their dominant hand. They were instructed to avoid mental counting or suchlike strategies, *e. g.* foot or hand tapping. The sessions always started with the r. t. task, which was followed by the time perception tasks in alternating order (reproduction/production or production/reproduction). In the beginning of the first experimental session, the tasks were explained to the subjects and practiced in three 'warm up' runs, with three trials for each task; these data were not analysed.

Data analyses and results

Of our interest are reproduction and production data and their relationships. In an earlier communication (Wackermann, Ehm, & Späti, 2003) we proposed the 'dual klepsydra model' (d. k. m.) of duration reproduction, yielding theoretical response curves. However, we have at present no comparable model for the production task. Therefore we focused on descriptive characterisation of response curves, using simplest phenomenological models like functional regressions. D. k. m.-based analyses of the t. r. t. data will be reported elsewhere.

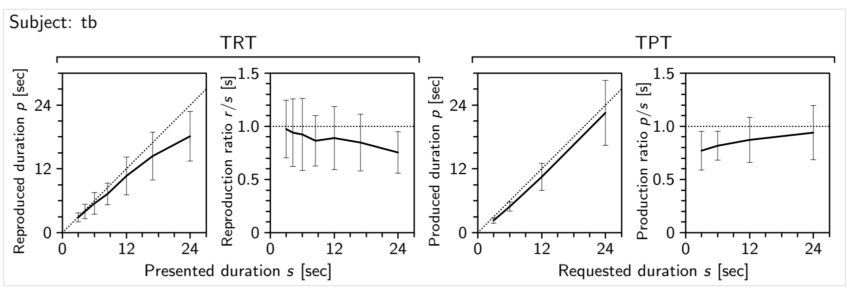


Fig. 2. Reproduction (TRT) and production (TPT) data for one individual subject, **tb**, collapsed across six sessions; shown are means ± 1 s. d. of the original response times, and of the response to stimulus ratios.

Data reduction

Data from a total of 59 experimental sessions were collected (number of sessions/subject varied from 4 to 6, mean = 4.9). Data from separate sessions were collapsed to single data sets for each individual subject, the data were inspected for extreme values, and outliers were removed.[4] An example of an individual data set reduced to means and standard deviations is given in Fig. 2.

Response variance

Dispersion of responses around individual means were, on the average, increasing with the stimulus magnitude s (retained duration in the t. r. t., or requested duration in the t. p. t.), as shown in Fig. 3. Assuming a power-function dependence of the standard deviation σ_s on s,

$$\sigma_s = \sigma_1 \cdot s^\gamma \,, \tag{1}$$

where σ_1 is an individual constant[5], exponents γ were estimated by linear regressions of $\log \widehat{\sigma}_s$ to $\log s$. Of theoretical interest are values $\gamma = \frac{1}{2}$ and $\gamma = 1$.

For the t. r. t., the estimates $\widehat{\gamma}$ ranged from 0.296 to 0.976, mean = 0.743, significantly higher than $\frac{1}{2}$ (Wilcoxon one-sample test, $T_- = 6$, $P < 0.01$), and all below 1. For the t. p. t., the estimates $\widehat{\gamma}$ ranged from 0.324 to 1.475, mean = 0.970, significantly higher than $\frac{1}{2}$ (Wilcoxon $T_- = 2$, $P < 0.01$), but not significantly different from 1 (Wilcoxon $T_- = 38$, *ns*). In pairwise between-tasks comparisons, the t. r. t. $\widehat{\gamma}$ exponents were mostly (9 of 12 subjects) *lower* than the t. p. t. exponents; the difference was marginally significant (Wilcoxon, $T_- = 18$, $P \approx 0.064$).

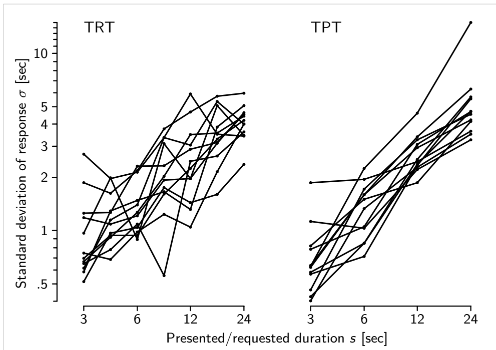


Fig. 3. Estimates of standard deviations $\widehat{\sigma}_s$ of response times plotted as functions of stimulus magnitude s in log–log co-ordinates. TRT = time reproduction task, TPT = time production task. Each curve represents one individual subject.

The seemingly 'smoother' appearance of the t. p. t. curves is due to the fact that s was varied across seven different values in the t. r. t., but only four different values in the t. p. t.. Note that some subjects show higher variance at s = 3 s than corresponds to a linear trend, $\log \sigma_s \approx c + \gamma \log s$.

Mean response curves

Characteristic forms of the response curves can be conveniently assessed by response-to-stimulus ratios q displayed as functions of stimulus magnitude s (Fig. 2). These functions can be modelled by a linear equation $q = a + b\,s$, which implies a quadratic regression model for the original response times,

$$r = a\,s + b\,s^2\,. \tag{2}$$

In eq. (2), $b = 0$ indicates linear response with a constant slope a; significant departures from $b = 0$ indicate non-linear response, with initial slope a at the origin $s = 0$. Coeffients a and b were estimated from the data by the weighted least squares method, using weighting factors $g_s \propto s^{0.75}$ for the t. r. t. data, or $g_s \propto s^1$ for the t. p. t. data, according to the typical γ values reported in the preceding paragraph.

For the t. r. t., the estimates of linear coefficients $\widehat{a}$ ranged from 0.687 to 1.59, mean = 1.069, not significantly different from 1 (Wilcoxon T_{-} = 28, *ns*). The estimates of quadratic coefficients $\widehat{b}$ ranged from -0.029 to -0.005 s^{-1}, mean = -0.015 s^{-1}. All $\widehat{b}$ coefficients were negative, for 11 of the 12 subjects highly significantly ($P <$ 0.005), indicating a negative curvature as a salient characteristic of the reproduction response curve.

For the t. p. t., the estimates $\widehat{a}$ ranged from 0.534 to 1.643, mean = 1.064; and the estimates $\widehat{b}$ ranged from -0.034 to $+0.013$ s^{-1}, mean = -0.001 s^{-1}, not significantly different from 0 (Wilcoxon T_{-} = 38, *ns*). We thus adopted the linear response model for the t. p. t., and re-evaluated the linear coefficients a_{o} for fixed $b = 0$; these estimates $\widehat{a}_{\mathrm{o}}$ ranged from 0.685 to 1.271, mean = 1.049, not significantly different from 1 (Wilcoxon T_{-} = 27, *ns*).

Finally, Spearman rank-correlations ρ between the regression coefficients were calculated (the subscripts 'r', 'p', and 'o' refer stand for 'reproduction', 'production', and linear model estimates of a's for 'production', respectively): $\rho(\widehat{a}_{\mathrm{p}}, \widehat{a}_{\mathrm{r}})$ = 0; $\rho(\widehat{a}_{\mathrm{p}}, \widehat{b}_{\mathrm{r}}) = -0.273$; $\rho(\widehat{b}_{\mathrm{p}}, \widehat{a}_{\mathrm{r}})$ = 0.385; $\rho(\widehat{b}_{\mathrm{p}}, \widehat{b}_{\mathrm{r}})$ = 0.084; $\rho(\widehat{a}_{\mathrm{o}}, \widehat{a}_{\mathrm{r}})$ = 0.238; $\rho(\widehat{a}_{\mathrm{o}}, \widehat{b}_{\mathrm{r}}) = -0.322$. None of these correlations was significantly different from 0.

Discussion and conclusions

Of particular interest is the dependence of response variance, σ_s^2, on the stimulus magnitude, s, since the exponents γ occurring in eq. (1) may provide a qualitative characterisation of the underlying processes. In both tasks, t. r. t. and t. p. t., the variance increased faster than it would be the case for a 'pure diffusion process', $\gamma = \frac{1}{2}$. Interestingly, γ's for duration production were apparently higher, on the average $\widehat{\gamma} \approx 1$, than γ's for duration reproduction.

In some subjects we have observed higher response variances at shortest durations, s = 3 s, than predicted by the power function, which otherwise holds good for longer durations (Fig. 3). This may indicate that different processes are activated in the domain of relatively short durations, $s \leq 3$ s. Indeed, times of 2—3 s are often reported as the upper limit of the 'subjective present' (Pöppel, 1978; 2004), while true experience of duration apparently develops only *beyond* that limit.[6] We are mainly interested in the metric of subjectively experienced duration, and we conclude that durations ≈3 s and shorter are better avoided.

The difference between the response curves forms characteristic for the t. r. t. and t. p. t., respectively, is also of importance. In the t. p. t., the produced times are generally linear functions of the requested time, with slopes roughly symmetrically distributed around $a = 1$. In the t. r. t., the reproduced times are non-linear functions of the stimulus duration, with initial slopes also symmetrically distributed around $a = 1$ but with a significant negative curvature, causing progressive shortening of response times with increasing stimulus magnitude. The effect is rather small (average $\widehat{b} \approx -1.5 \times 10^{-2}\,\mathrm{s}^{-1}$) and becomes evident only with prolonged stimulus durations. We already have pointed out the importance of this effect for the construction of theoretical models of internal time representation (Wackermann, Ehm, & Späti, 2003), but our work relied on data reported in literature; now we are in possession of an experimental database against which further modelling developments can be tested.

We did not find any significant correlations between parameters of response curves obtained in the reproduction and production tasks. However, the relationships between duration reproduction and production may be too complex to yield a clear picture from a small-size sample ($N = 12$). The present analyses assessed only mean tendencies of individual subjects and focused on across-subjects correlations; further analyses may or may not reveal correlations between reproduction and production at the intra-individual level.

The lack of an obvious relation between reproduction and production, the qualitative difference between the typical forms of the response curves, and the difference between characteristic exponents of the response variance—all these findings lead back to the questions raised in the introduction: what do these different methods measure?

It is often assumed that numerical estimation, and the inverse task, production, reveal the functions mapping the perceived physical duration to a hypothetical internal representation, 'subjective duration', or *vice versa*. Reproduction would then be conceivable as a composition of two processes,

$$\underbrace{\text{estimation} \rightarrow \text{(internal state)} \rightarrow \text{production}}_{\text{reproduction}}\,. \tag{3}$$

But, assume initially that the 'estimation' component in (3) is a deterministic function, mapping the perceived duration onto the internal state with zero error. Then the response variance would be entirely due to the error of the 'production' component; the characteristic exponent (c. e.) γ_{r} of the reproduction response variance should thus be identical to the c. e. γ_{p} of the production response variance. More realistically, the 'estimation' component may also be subject to error variance, with a c. e., say, γ_{e}. The c. e. of the composite process should *never be lower* than the c. e. of any one of its components; however, in our data we see predominantly $\widehat{\gamma_{\mathrm{r}}} < \widehat{\gamma_{\mathrm{p}}}$. We thus conclude that the composite model of reproduction in the form (3) may not be correct. This would corroborate our concept of time reproduction as the very elementary process upon which other forms of temporal cognition rely.

Notes

[1] The experiments were conducted by J. S. for his diploma thesis (University of Freiburg, Department of Biology), under advise and supervision by J. W. and W. E.

[2] Unlike chronobiological studies carried out in controlled environments, our experiments were arranged in the participants' natural environments, in parallel with their usual daily activities. To assess the individual diurnal cycle of the subjects, their sleep/waking times were protocolled, and body temperature was regularly measured. These additional data, as well as the r. t. data, are not reported here.

[4] A simple 'data peeling' algorithm was used for outliers detection. Given a data set $x_1, \ldots, x_n$, a x_j was classified as an outlier if $|x_j - m|/\sqrt{v} > c$, where

$$m = (n-1)^{-1} \sum_{i \neq j} x_i\,, \quad v = (n-2)^{-1} \sum_{i \neq j} (x_i - m)^2,$$

are the arithmetic mean and variance of the sample with element x_j removed, and c is a chosen constant. All data elements classified as outliers were removed from the data set, and the procedure was iterated with the reduced data set until no more outliers were detected. Exclusion criterion $c = 3$ was used for all data sets, resulting in an average data loss of 3.8% (t. r. t.), 2.5% (t. p. t.), and 9.7% (r. t.).

[5] Nominally, σ_1 stands for the standard deviation of the response to the retained/requested duration $s = 1$ s. But such an extrapolation is not justified, because the (re)productive behaviour may be qualitatively different in the domain of very short durations—*cf.* also note 6 and the corresponding passage in the Discussion.

[6] H. Woodrow (1930) used in his experiments stimuli of durations from 0.2 s through 30 s, and observed that the subjects "seemed all to be of the opinion that, if there is any experience which deserves to be called the perception of time, that such experience begins to become definite only at 1.5 to 6 sec, and continues as much the same quality of experience throughout all the longer intervals." (p. 494). —The phenomenological evidence for a lower bound ('horizon') of experienced duration is commonly ignored in contemporary research, which is biased towards the second and sub-second range. It is questionable whether results obtained in the sub-second domain are representative for human experience of time *in toto*.

References

Bindra, D. & Waksberg, H. (1956). Methods and terminology in studies of time estimation. *Psychological Bulletin*, **53**, 155–159.

Church, R. M. (1984). Properties of the internal clock. In J. Gibbon & L. Allan (Eds.) *Timing and Time Perception*. *Annals of the New York Academy of Sciences*, **423**, 566–582.

Eisler, A. D. (2003). The human sense of time: biological, cognitive and cultural considerations. In R. Buccheri, M. Saniga, & W. M. Stuckey (Eds.) *The Nature of Time: Geometry, Physics and Perception*, Dordrecht: Kluwer, pp. 5–18.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, **127**, 22–44.

Pöppel, E. (1978). Time perception. In R. Held, H. W. Leibowitz, H.-L. Teuber (Eds.) *Handbook of Sensory Physiology; Volume 8: Perception*, Berlin: Springer, pp. 713–729.

Pöppel, E. (2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis*, **64**, 295–301.

Pöppel, E. & Giedke, H. (1970). Diurnal variation of time perception. *Psychologische Forschung*, **34**, 182–198.

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen.* Tübingen: Laupp.

Wackermann, J. (2005). On internal time measures and properties of the klepsydraic clock. In J. Monahan, J. Townsend, & S. Sheffert (Eds.), *Fechner Day 2005*, Traverse City: International Society for Psychophysics, pp. 000–000.

Wackermann, J., Ehm, W., & Späti, J. (2003). The 'klepsydra model' of internal time representation. In B. Berglund & E. Borg (Eds.) *Fechner Day 2003*, Stockholm: International Society for Psychophysics, pp. 331–336.

Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, **13**, 473–499.

Woodrow, H. (1951). Time perception. In S. S. Stevens (Ed.) *Handbook of Experimental Psychology*, New York: Wiley, pp. 1224–1236.

Zakay, D. (1990). The evasive art of subjective time measurement: some methodological dilemmas. In R. A. Block (Ed.) *Cognitive Models of Psychological Time*, Hillsdale NJ: Lawrence Erlbaum, pp. 59–84.

A META-ANALYSIS OF THE EFFECTS OF CONTEXT ON JUDGMENTS OF SIZE AND DISTANCE

Mark Wagner

Department of Psychology, Wagner College, Staten Island, NY 10301, mwagner@wagner.edu

Abstract

This paper presents the results of a meta-analysis on the space perception literature. Conditions that significantly influence size and distance judgments and the degree to which they alter the exponent of the power function are examined. It also looks at how well the power function fits spatial data and under which circumstances the fits are particularly good or particularly poor. Finally, multivariate analyses are preformed.

Perhaps the first meta-analytic review of the spatial literature may be attributed to Baird (1970). Baird reported 69 exponents from 14 studies on judgments of length, area, and volume. Baird found that exponents for length are largest when the standard is in the middle of the range. He also discussed the effects of average stimulus size, stimulus range, and instructions, but believed that these variables had little influence of length judgments. Overall, Baird believed the average exponents for length, area, and volume averaged about 1.0, .8, and .6, respectively.

Two meta-analytic reviews of distance perception appeared in 1985. DaSilva (1985) reported 76 exponents from 32 studies on egocentric distance estimation, where egocentric distance refers to the distance from the observer to an object. DaSilva excluded exocentric distance judgments, which refers to the length or distance between two endpoints that are both located away from the observer. Based on this analysis and a number of his own experiments, DaSilva concluded that the typical exponent for egocentric distance estimation was about .90. He also found that magnitude estimation produces lower exponents that ratio estimation or fractionation, that increasing stimulus range produces lower exponents, and that the exponents for individual subjects are quite variable but less than 1.0 about 78% of the time.

Wiest and Bell (1985) analyzed 70 exponents taken from 25 studies on distance estimation. Their main finding concerned performance across perceptual, memory, and inference conditions. To Wiest and Bell, judgments are perceptual when the judged stimuli are available to the subject throughout the judgment process; memory judgments occur when stimuli are presented perceptually to the subject at one time but judgments are made at a later time when the stimuli are no longer available, and inference judgments occur when knowledge about stimulus layout is acquired across time such as in cognitive mapping. Wiest and Bell found that the average exponents were 1.08, .91, and .7 for the perceptual, memory, and inference conditions respectively. They also found that larger stimulus ranges are associated with smaller exponents and that judgments collected outdoors tend to produce smaller exponents than those collected indoors.

A great deal of spatial perception research has been reported in the decades following these prior reviews. The present review updates and greatly expands on these reviews, because it is based on over seven times as much information as the most extensive previous work.

Methodology

The initial list of works to include in the analysis came from a computer search for English-language articles on length, distance, size, area, volume, and space perception that included mention of power functions or exponents. The list was then supplemented by adding references from the previous meta-analytic reviews (Baird, 1970; DaSilva, 1985; Wiest & Bell, 1985) and other articles.

In the end, the analysis includes 104 space perception articles and a total of 530 power function exponents. The later number is greater than the former because most articles reported exponents for multiple experimental conditions. For each study, 13 variables were recorded as listed in Table 1. The results for area, volume, and angle judgments are reported in *The Geometries of Visual Space* (Wagner, 2005). The present article focuses on size and distance judgments, however, a more complete analysis of this data are also found in Wagner (2005).

The effects of each variable were analyzed three ways. First, an analysis of the entire data set will be discussed. Second, the analysis will be repeated for perceptual data alone, excluding inference and memory conditions. Third, the analysis will be refined even further to focus on perceptual data collected using magnitude estimation or ratio estimation. (Two methods which Wiest and Bell (1985) and others argue are equivalent.)

The data are presented in a variety of formats in recognition of the dangers involved in meta-analytic research. Data collected under memory and inference conditions may not be equivalent to perceptual data. Similarly, certain judgment methods such as mapping, fractionation, or triangulation place physical constraints on the judgment process that may render them non-equivalent to numeric techniques like magnitude and ratio estimation.

There were a total of 362 exponents for distance estimates that entered into the analysis. The perceptual data set had 257 exponents, while the magnitude/ratio estimation data set had 182 exponents. Table 1 summarizes the effects of different contextual variables on judgments of size and distance. I will discuss the effects of each variable in turn.

Overall. In general, exponents for distance estimation are very close to 1.0. For both perceptual data sets, it appears that the average exponent is slightly larger greater than 1.0, and the data are well-fit by a power function with coefficients of determination averaging about .95. The total data set shows a lower average exponent and a lower coefficient of determination due to the influence of memory and inference conditions.

Age. Only a small number of studies reported the average age of the subjects involved. Typically, studies either reported an age range or described the population from which subjects were drawn—such as undergraduates. I used this information to estimate the average age of the subjects. In truth, there was little variation in age because 84% of all studies were based on undergraduate students.

Based on the total data set, exponents appeared to significantly increase with age, $r = .15$, $p < .05$, but this trend is not significant for the other two data sets. Coefficients of determination decline with age for the total data set, $r = -.19$, $p < .05$, and for perceptual judgments, $r = -.28$, $p < .05$.

Cue conditions. Some studies limit perceptual exploration and/or cues to distance more than others. Full-cue conditions allow relatively complete layout information, while reduced-cue conditions limit perceptual information in some way. In the present analysis, most studies (81.2%) were classified as full cue.

Exponents were significantly greater under full-cue conditions for all data combined and for perceptual magnitude estimates. In truth, the effects of cue conditions may be greater than the summary data would make it seem because 60.4% of the studies classified as "reduced-cue" only limited information by fixing head position with a bite bar. This attempt at increased experimental control may also explain why the coefficient of determination was significantly greater under reduced conditions.

Egocentric vs. exocentric. There were no significant differences in the exponent between egocentric and exocentric distance judgments. Coefficients of determination, however, were significantly higher for egocentric judgments for all three data sets.

Table 1

Mean Exponent and Coefficient of Determination (R^2) as a Function of Recorded Variables for Judgments of Distance Calculated from All Data, Perception Data, and Magnitude or Ratio Estimation Perception Data

Variable	Value of Variable	*Exponent* All	*Exponent* Per.	*Exponent* Mag. Est.	*Coefficient of Determination* All	*Coefficient of Determination* Per.	*Coefficient of Determination* Mag.Est.
Overall		0.96	1.02	1.04	0.91	0.95	0.96
Age	Pre Under.	0.93	0.92	1.05	0.97	0.97	0.99
	Undergrad.	0.95	1.02	1.02	0.91	0.97	0.96
	Post Under.	1.04	1.04	1.05	0.91	0.91	0.95
Cue Conditions	Full	1.00	1.02	1.04	0.92	0.94	0.95
	Reduced	0.99	0.99	1.01	0.98	0.98	0.99
Ego vs. Exocentric	Egocentric	0.99	1.01	1.06	0.96	0.98	0.99
	Exocentric	0.95	1.03	1.01	0.88	0.91	0.92
Immediacy	Perception	1.02		1.04	0.96		0.96
	Memory	0.87		0.90	0.87		0.85
	Inference	0.77		0.76	0.82		0.81
Inside or Outside	Inside	1.02	1.05	1.04	0.91	0.96	0.95
	Outside	0.89	0.97	1.03	0.91	0.95	0.97
Instructions	Objective	0.94	1.03	1.08	0.90	0.97	0.96
	Neutral	0.93	1.00	1.03	0.90	0.94	0.92
	Apparent	1.02	1.03	1.04	0.96	0.95	0.97
	Road Path	0.70			0.82		
Method	Fractionation	0.89	0.90		0.97	0.97	
	Ratio Est.	0.82	1.02	1.02			
	Mag. Est.	0.98	1.04	1.04	0.91	0.96	0.96
	Mapping	0.76	0.96		0.88		
	Production	1.02	1.06		0.92	0.99	
	Triangulation	0.96	0.96				
Number of Subjects	≤ 15	1.00	1.03	1.03	0.92	0.94	0.95
	> 15	0.90	1.01	1.04	0.91	0.97	0.97
Standard Used	No	0.96	1.03	1.05	0.93	0.98	0.98
	Yes	0.96	1.00	1.00	0.88	0.92	0.93
Standard Size	Small	0.94	1.02	1.02	0.84	0.82	0.86
	Midrange	0.96	0.94	0.96	0.93	0.92	0.93
	Large	1.04	1.03	1.03	0.98	0.98	0.98
Stimulus Orientation	Horizontal	1.02	1.01	1.02	0.95	0.93	0.86
	Vertical	1.04	1.04	1.04	0.91	0.91	0.91
	In depth	1.01	1.04	1.09	0.95	0.98	0.98
Log Stimulus Range	< 1	1.06	1.06	1.10	0.97	0.97	0.96
	1 ≤ x ≤ 1.5	1.06	1.06	1.08	0.94	0.94	0.96
	> 1.5	0.86	0.89	0.91	0.98	0.97	0.98
Stimulus Size	< 1 m	1.04	1.04	1.02	0.95	0.95	0.94
	1m < x < 10m	1.10	1.09	1.11	0.99	0.99	0.99
	> 10m	0.88	0.93	0.99	0.90	0.94	0.96

Immediacy. For want of a better word, immediacy here refers to the effects of perceptual, memory, and inference conditions as defined by Wiest and Bell (1985). The stimulus is most immediately available in perceptual conditions and never directly available under inference conditions.

The exponent is highest for perceptual conditions and lowest in the inference condition. In addition, coefficients of determination were highest for perceptual conditions and lowest for inference conditions. These differences in exponents and coefficients of determination between all three conditions are statistically significant at the .001 level according to Duncan follow up tests.

Wiest & Bell (1985) restricted their analysis to magnitude and ratio estimates. To make the present analysis equivalent, Table 1 also reports mean exponents for perceptual, memory, and inference conditions based exclusively on magnitude and ratio estimates. (This is the only time the third column of the table includes memory and inference conditions.) The exponents reported here are similar to those reported by Wiest and Bell; however, they deviate a bit less from 1.0 than in their report. Once again, coefficients of determination are highest for perceptual conditions and lowest for inference conditions.

Inside vs. outside. Experiments conducted indoors produce significantly higher exponents than those conducted outside for the two larger data sets. However, this difference is not significant for perceptual data collected using only magnitude/ratio estimation. Coefficients of determination do not differ significantly between inside and outside studies.

Instructions. Studies differ in the way they describe the judgment task to subjects. In the present analysis, instructions were classified into four types. Apparent size instructions were the most commonly employed (39.2% of studies). These instructions ask subjects to judged how the distance "looks", "appears", or "seems to be" subjectively. Objective instructions, which explicitly emphasize physical accuracy, are less commonly employed (24.3% of studies). A third categorization used here was a neutral category, that neither emphasized physical accuracy nor how things appear subjectively, but rather simply asked subjects to judge the distance between two points. This categorization describes 34.3% of the studies. Finally, a relatively small number of studies (2.2%) asked subjects to judge the distance from one place to another, not as the crow flies, but according to the length of the route one would need to take to drive from one place to another.

Instructions appeared to have inconsistent effects on the exponent. For the complete data set, apparent instructions produced significantly higher exponents than objective or neutral instructions. For perceptual magnitude/ratio estimates, objective instructions tended to result in higher exponents than either apparent or neutral instructions although this trend is not significant. Road path length produced consistently and significantly lower exponents and coefficients of determination than other instruction types.

Method of judgment. Analysis of variance indicates that the exponent differs significantly as a function of method used to collect judgments for the total data set at the .001 level. Duncan follow up tests indicate that mapping, complete ratio estimation, and fractionation exponents are significantly smaller than exponents based on magnitude estimation and production. Exponents also differ significantly (at the .001 level) as a function of method for the perceptual data set. Here, Duncan follow up tests reveal that the exponents for magnitude estimation, complete ratio estimation, and production are all significantly greater than those produced by fractionation. The magnitude/ratio estimation data set found no significant difference between magnitude and ratio estimation. There was too little data collected employing category estimation or absolute judgment to make meaningful statistical statements, but such as there is indicates that category estimation exponents seem to be low (.90) while absolute judgment exponents seem to be very high (2.44). Coefficients of determination did not differ significantly as a function of method.

Number of subjects. The number of subjects in the experiment proved to be significantly negatively correlated with the size of the exponent for the total data set, $r = -.30$, $p < .001$. The most probable explanation of this relationship is that memory studies typically had many more subjects than perceptual ones. Because memory exponents are much lower than perceptual exponents, this would explain why exponents went down with the number of subjects under memory conditions. When memory is factored out as in the two perceptual data sets, the correlation between the number of subjects and the exponent disappears.

Standard. Although no difference is found with the total data set, studies using a standard had significantly lower exponents and coefficients of determination than studies that used no standard for perceptual data collected using magnitude or ratio estimation. Using a standard gave rise to more accurate judgments with the exponent averaging precisely 1.0. It would appear that putting the standard in the middle of the stimulus range gives rise to lower perceptual exponents than placing it at either extreme, but this trend is not significant.

Stimulus orientation. The orientation of the stimulus did not significantly influence the exponent for any of the data sets. However, for both of the perceptual data sets, orientation significant influenced the coefficient of determination. In both data sets, Duncan follow up tests show that in-depth oriented stimuli gave rise to significantly higher R^2 values than frontally presented stimuli in a vertical orientation. For the magnitude/ratio estimates, in-depth stimuli were also associated with higher R^2 values than frontally presented stimuli oriented horizontally.

Stimulus range. Teghtsoonian (1971, 1973) proposed that the exponent was closely related to the range of stimuli presented to the subject. Large stimulus ranges should theoretically produce consistently smaller exponents. In the present study, the stimulus range was determined dividing the maximum stimulus presented to the subject by the smallest. To be consistent with Teghtsoonian's work, I then took the logarithm of this ratio.

Stimulus range proved to be one of the most powerful predictors of the exponent. The exponent was negatively correlated with log stimulus range at the .001 level for all three data sets ($r = -.45$ for all data, $r = -.38$ for perception data, $r = -.44$ for perceptual magnitude estimation data). The coefficient of determination was not significantly correlated with stimulus range for any of the data sets.

Stimulus size. For the total data set, stimulus size—defined as the midpoint of the stimuli presented to the subject—was significantly negatively correlated with the size of the exponent, $r = -.16$, $p < .01$. Once again, this correlation probably arises because inference conditions, which are associated with lower exponents, often use large-scale environments to test cognitive mapping knowledge. When memory and inference conditions are factored out, as in the two perceptual data sets, the correlation between stimulus size and the exponent is no longer significant.

Multivariate Analyses

Of course, there are certain statistical dangers associated with conducting a large series of univariate significance tests. First of all, the more tests one conducts, the higher the likelihood that some of the significant findings arise by chance (Type I error). Secondly, it is possible that real trends in the data may be produced (or obscured) by the effects of secondary factors that are accidentally associated with the variable due to the non-random nature of the data collection process. Because some combinations of conditions may occur more frequently than others, it is possible for variables to be correlated with one another. Thus, the apparent influence (or lack of influence) of one variable on the exponent might really be due to it being correlated with another recorded variable.

To overcome these deficiencies, I performed a series of linear multivariate regression analyses. The exponent was the outcome variable and all variables that displayed any significant univariate associations with the exponent served as predictor variables. These predictor variables included age, cue conditions, immediacy, inside vs. outside location, instructions, method, number of subjects, standard presence, and log stimulus range. Mean stimulus size was not included as a predictor because its association with the exponent was non-linear. Variables with no significant univariate association with the exponent were excluded to limit the multivariate model to a reasonable size.

Categorical variables were recoded as dummy variables where "1" represented the presence of a factor and "0" represented the absence of a factor. If more than two levels existed for a categorical factor, a series of dummy variables were used to represent the information. For example, instructions were broken down into two variables, one for the presence of objective instructions and one for the presence of apparent instructions.

For the total data set, four factors proved to be significantly associated with the exponent in the multivariate analysis: inference conditions ($t = 3.70$, $p < .001$), the presence of a standard ($t = 2.64$, $p < .01$), judgment method ($t = 2.84$, $p < .005$), and log stimulus range ($t = 9.54$, $p < .0001$). Once the significant variables were determined, a second regression analysis was performed employing only these four significant variables to determine the best fitting equation to predict the exponent. This equation accounted for 42.03% of the variance in the exponent.

The results of the second regression equation can be substituted into a power function to yield the following general equation to predict distance judgments:

$$J = \lambda D^{[.083(\text{mag}) - .058(\text{stan}) - .210(\text{inf}) - .167(\log(\text{max/min})) + 1.172]} \quad (1)$$

where J is the subject's judgment for distance, λ is a scaling constant, D is the physical distance, "mag" is code "1" if the method is magnitude or ratio estimation and "0" otherwise, "stan" is coded "1" if a standard is used and "0" otherwise, "inf" is code "1" for inference stimuli and "0" for memory and perceptual stimuli, and log(max/min) refers to the base 10 logarithm of the ratio of the largest stimulus used in the experiment to the smallest.

For the perceptual data set, three (of the eight remaining) variables proved to be significantly related to the exponent in the multivariate regression analysis. Consistent with the total data set, the three variables that were significantly related to the exponent were method ($t = 2.75$, $p < .01$), presence of a standard ($t = 2.40$, $p < .05$), log stimulus ratio ($t = 8.17$, $p < .0001$). (Of course, the inference condition could not be an element in the perceptual equation because inference data was factored out of this data set.) When a regression equation was generated based on these three factors it accounted for 29.46% of the data. Substituting this linear equation into the power function yields the following general equation to predict perceptual distance judgments:

$$J = \lambda D^{[.084(\text{mag}) - .055(\text{stan}) - .166(\log(\text{max/min})) + 1.261]} \quad (2)$$

Finally, for the perceptual data set focusing on magnitude and ratio judgments, only two factors were significantly related to the exponent in the multiple regression equation, presence of a standard ($t = 3.02$, $p < .01$) and log stimulus ratio ($t = 6.25$, $p < .0001$). This regression equation accounted for 23.48 % of the variance in the exponent. Substituting this equation into the power function yields the following equation to predict distance judgments:

$$J = \lambda D^{[-.073(\text{stan}) - .136(\log(\text{max/min})) + 1.223]} \quad (3)$$

References

Baird, J. C. (1970). *Psychophysical analysis of visual space*. Oxford: Pergamon.

Da Silva, J. A. (1985). Scales for perceived egocentric distance in a large open field: Comparison of three psychophysical methods. *American Journal of Psychology*, *98*, 119–144.

Teghtsoonian, R. (1971). On the exponent of Stevens' law and the constant in Ekman's law. *Psychological Review*, *78*, 71–80.

Teghtsoonian, R. (1973). Range effects in psychophysical scaling and a revision of Stevens' law. *American Journal of Psychology*, *86*, 3–27.

Wagner, M. (2005). *The geometries of visual space*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wiest, W. M. & Bell, B. (1985). Steven's exponent for psychophysical scaling of perceived, remembered, and inferred distance. *Psychological Bulletin*, *98*, 457–470.

Loudness Adaptation: Peripheral Implications in the Perceived Loudness Decline of Constant Intensity Tones

Ernest M. Weiler, Ph.D.,David E. Sandman, BA
Hongwei Dou, PhD., Kathleen Cross, MA, Thomas Goldman, Ph.D.
University of Cincinnati, Cincinnati, Ohio
Contact Ernest M. Weiler, ML#379
University of Cincinnati, Cincinnati, Ohio, 45267.

Abstract

This paper summarizes evidence that Loudness Adaptation measured by the monaural Ipsilateral Comparison Paradigm (ICP) has a clear foundation in the peripheral auditory system, rather than originating in central "induced' effects. The argument proceeds by recalling the association of ICP loudness adaptation with three established peripheral phenomena. First we note our previous findings of a simple association of the ICP with classic Tone Decay (TD). Secondly, we recall our findings that the ICP adaptation function changes in individuals with cochlear (peripheral) hearing pathology. Thirdly, we replicate findings by Collet & various colleagues showing that suppression of Oto-acoustic Emissions (OAE) is associated with TD adaptation, and reiterate our findings that ICP adaptation is also associated with OAE suppression, as well as being associated with classic TD. Lastly we consider the puzzling lack of simple adaptation (SA) at medium acoustic intensities for those tones most important to speech perception. We conclude the monaural ICP Loudness Adaptation originates in the auditory periphery, even though revealed by means of Psychophysical (central) judgments.

When Derbyshire & Davis (1935) observed that stimulation of individual auditory neurons produced a decline in activity over several minutes there was no doubt that they were observing peripheral adaptation. However, the issue is not so clear when psycho-acoustical judgments are used to evaluate the subjective *loudness* and *loudness adaptation* (decline) of a stimulus tone. Can evidence be found to support a direct relationship between psycho-acoustical loudness adaptation and peripheral events? Weiler, Gold, Sandman & Warm (1992) analyzed a number of measures of loudness adaptation to a principle components factor analysis. They found there were monaural, apparently peripheral components of variance in the measures used, the peripheral influence could only be seen by statisticaly removing binaural/interaural effects.

Peripheral effects and Loudness adaptation

In the search for a connection between psychoacoustical judgements of loudness adaptation and peripheral adaptation at the sensory level, we have examined the following phenomena: 1) classic monaural near-threshold *tone decay* adaptation (after Carhart, 1957), 2) *simple loudness adaptation* based on judgments of a monaural steady tone (Scharf, 1983), 3) the monaural *Ipsilateral Comparison Paradigm* adaptation (after Weiler, Sandman & Pederson, 1981, and Dange, Warm, Weiler & Dember, 1993), 4) *suppression of otoacoustic emissions* (OAE), which are measures from the cochlea (after Kemp, 1978), and 5) relationships between diagnosed *cochlear (peripheral) hearing loss* with magnitude estimated loudness adaptation, but not with reaction time adaptation (Goldman, Weiler, & Davis (1981).

Monaural near-threshold tone decay (point 1 above) and cochlear hearing loss (point 5) are strongly associated and tone decay has long been an accepted diagnostic indicator of sensory hearing loss in the cochlea, since Carhart (1957). This paper will discuss a series of studies showing that monaural loudness adaptation well above threshold, as in points 2,3, and 4, may also be associated with peripheral hearing loss or suppression of otoacoustic emissions, from the healthy cochlea.

Methods

1) TD, classic monaural tone decay is used as a gold standard of peripheral/sensory loudness adaptation in this study. TD, has been a durable method discovered in normal hearing individuals (Ward, 1973). However, those with a cochlear hearing loss show much more extreme tone decay effects than normals. The magnitude of adaptation is determined by the sum of dB increments necessary to maintain the continuous stimulus at a threshold level. Note that the present study only investigates TD adaptation in normal hearing individuals.

2) SA, the method of simple loudness adaptation (Scharf, 1983) has also been called the method of monaural successive magnitude estimation of loudness adaptation. The method assesses the simple judged decline in loudness of a tone presented at a fixed intensity. The listener is required to make numerical magnitude estimates of the loudness of the steady tone successive time intervals. Adaptation is said to occur when the loudness magnitude of the final estimate is smaller than that made at the onset of the tone.

3) ICP, the ipsilateral Comparison Paradigm procedure (Weiler, Sandman and Pederson, 1981; Dange, Warm, Weiler and Dember, 1993; Tannen et al., 2001; Sandman, Weiler, and Pederson, 1982; Jones, et al., 2003; and Weiler et al., 2000), reveals loudness adaptation at a wide range of supra-threshold frequencies and intensities, which do not show a decline by the SA procedure. Like the method of simple loudness adaptation, numerical magnitude estimation of loudness is used to assess adaptation of a base tone of specified intensity. Unlike the SA procedure, the ICP applies periodic increments of 10 dB (10 sec, every 30-sec) to the intensity of the base tone which serves generally as the adapting tone. The intensity modulation of the base tone was thought to serve as a reference tone. Listeners make loudness judgements in the form of magnitude estimates of both the adapting and the reference tones. Adaptation is determined based on the difference between the last and the first estimates of the loudness of the baseline.

4) Oto-acoustic emissions (OAEs) serve as the objective measure of peripheral activity in the present study. OAEs are acoustic energy emitted by the cochlea that can be measured in the outer ear canal (Kemp, 1978). They are generated by the motile activity of the outer hair cells and represent the active mechanism within the cochlea (Brownnel, 1990; Dallos, P., and Evans, B.N., 1995). The evoked emissions are elicited

by sounds presented to the ear. The present study used evoked emissions defined as transient evoked OAEs (TEOAEs). TEOAEs are recorded following acoustic stimulation of clicks or tone pips to the ear.

5) Relationships between adaptation and hearing loss have been investigated for both TD and ICP loudness adaptation. Once again, Carhart (1957) serves as a foundation for use of abnormal TD and those with cochlear loss. Janson, et al. (1996), compared normal listeners to those with high frequency cochlear hearing loss (peripheral/sensory) and found abnormal ICP adaptation for the impaired at 1000 Hz where their thresholds were normal.

Probing Peripheral Monaural Adaptation

Some researchers suggest that the negative feedback from cochlear efferent fibers may be a way to access auditory adaptation (Leibrandt, 1965; Collet et al., 1992). The degree of adaptation is thought to be related to the strength of the function of the medial olivocochlear system (MOC) (Collet et al., 1992; Micheyl, Carbonnel and Collet, 1995). The MOC system is composed of cochlear efferent neurons arising from the medial nuclei of the superior olivary complex; these neurons travel mainly to the contralateral outer hair cells (OHCs). It has been demonstrated that MOC neurons can be activated ipsilaterally, contralaterally, and bilaterally with auditory stimulation of sufficient duration (Berlin, Hood, et al., 1995). They appear to function as a feedback loop causing some type of regulation or inhibition on peripheral auditory activities (Galambos, 1956; Buno, 1978). Recently, some investigators postulated that the inhibitory effect of MOC neurons on the auditory periphery was through their modulation of outer hair cell function (Mountain, 1980; Collet et al., 1990; Berlin, Hood et al., 1995). This assumption was supported by the evidence that activation of efferent neurons resulted in changes in otoacoustic emission responses, including amplitude, frequency, and temporal variations (Mountain, 1980; Collet, et al., 1990; Berlin, Hood, et al., 1995).

Loudness Adaptation Correlated with Suppression of OAEs

Following the logic above, Collet and colleagues have shown that the classic tone decay adaptation correlated with suppression of otoacoustic emissions (OAE) within the cochlea. In particular, Collet, Veuillet, Micheyl, Amabile, and Morgon (1992) found a statistically significant correlation of $r = -0.60$, while Micheyl, Morlet, Giraud, Collet and Morgan (1995) found significant correlations ranging from $r = -0.43$ to -0.70 depending on conditions. Inspired by these findings, Dou and colleagues, (Weiler, Dou, Tannen, Sandman, Dember, and Warm, 1998) compared ICP loudness adaptation to suppression of OAE's and found a significant relationship ($r = -0.36$) for 41 normal hearing adults. A subsequent study with 75 normal hearing adults, Dou (1998, 2000), showed significant correlations at 60 dB for ICP adaptation with suppression ($r = -0.27$), TD with suppression ($r = -0.34$) and SA with suppression ($r = -0.28$). Furthermore, ICP and TD adaptation showed a significant correlation ($r = + 0.26$), but neither correlated with SA. We conclude that OAE suppression is useful to study of peripheral auditory adaptation, and the phenomena repeatedly correlates with TD and ICP loudness adaptation.

Clinical evidence supporting the peripheral nature of TD and ICP adaptation

We refer to further evidence that both TD and ICP adaptation are altered for those with a peripheral / sensory cochlear hearing loss. Carhart (1957) established this for TD, Janson, et al. (1996) found the cochlear loss changed the ICP, and we have previously mentioned that we found significant correlations between TD and ICP adaptation.

CONCLUSION

Since loudness adaptation from both Tone decay (TD) and the Ipsilateral Comparison Paradigm (ICP) procedures are correlated with suppression of OAE effects, and with each other, it is concluded that ICP adaptation also has a peripheral basis. The clinical evidence for those with cochlear loss supports this conclusion.

REFERENCES

Berlin, C.L., Hood, L.J., Jurley, A., Wen, H., and Kemp, D.T. (1995). Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise. Hearing Research, Vol. 87, 96-103.

Brownell, W.E. (1990). Outer hair cell electromotility and otoacoustic emissions. Ear and Hearing, Vol. 11, 82-92.

Buno, W. (1978). Auditory nerve fiber activity influenced by contralateral ear sound stimulation. Experimental Neurology, Vol. 59, 62-74.

Carhart, R. (1957). Clinical determination of abnormal auditory adaptation. Archives of Otolaryngology. 65, 32-39.

Collet, L., Kemp, D.T., Veuillet, E., Duclauz, R., Moulin, A., & Morgon, A. (1990). Effect of contralateral auditory stimuli on active cochlear micro-mechanical properties in human subjects. Hearing Research, 43, 251-262.

Collet, L., Veuillet, E., Micheyl, C., Amabile, J.C., & Morgon, A. (1992). Involvement of medial olivocochlear system in loudness adaptation. Advances in Biosciences, 83, 297-303.

Dallos, P., and Evans, B.N. (1995). High frequency motility of outer hair cells and the cochlear amplifier. Science, Vol. 267, 2006-2009.

Dange, A., Warm, J.S., Weiler, E.M., and Dember, W.N. (1993). Loudness Adaptation: Resolution of a psychophysical engima. Journal of General Psychology, 120, 217-243.

Derbyshire, A.J. & Davis, H. (1935). The action potentials of the auditory nerve. American Journal of Physiology, 113, 476-504.

Dou,H. (1998). Loudness adaptation and transient evoked otoacustic emissions: investigating the potential neurophysiological correlates of loudness adaptation. Ph.D.

dissertation in the dept.of Communication Sciences and Disorders of College of Allied Health Sciences, University of Cincinnati, Cincinnati, Ohio.

Dou, H., Weiler, E.M., Maguire, T., Boyce, S., Kretschmer, L. Shiau, W, and Sandman, D.E. (2000). Is speech perception modulating loudness adaptation? Association for Research in Otolaryngology. Abstract, 23 meeting, St. Petersburg Beach, Fla.

Galambos, R. (1956). Suppression of auditory activity by stimulation of efferent fibers to the cochlea. Journal of Neurophysiology, Vol. 19, 424-437.

Goldman, T, Weiler, E.M., and Davis, J.M. (1981). Perstimulatory adaptation as measured by reaction time: Asymmetry between 400 and 500 kc/s. Journal of Auditory Research, 21, 13-16.

Janson, J.L., Weiler, E.M., Sandman, D.E., Kretschmeter, L.L., Nelson, T., & Stromberg, E. (1996). Abnormalities of ICP loudness adaptation at 1000 Hz associated with higher frequency (cochlear) loss. British Journal of Audiology, 30, 37-43.

Jones, K.S., Weiler, E.M., Warm, J.S., Dember, W.N. and Sandman, D.E. (2003). Loudness adaptation: fact or artifact? Journal of General Psychology, Vol. 130, 341-358.

Kemp, D.T. (1978). Stimulated acoustic emissions from within the human auditory system. Journal of the Acoustical Society of America, 64, 1386-1397.

Leibrandt,C.C. (1965). The significance of the olivocochlear bundle for the adaptation mechanism of the inner ear. Acta Oto-Laryngologica, Vol. 59, 124-131.

Micheyl. C., Morlet T., Giraud, A.L., Collet, L., and Morgon, A. Contralateral Suppression of Evoked Otoacoustic Emissions and Detection of a Multi-Tone Complex in Noise. Acta Otolaryngol (Stockh) 1995; 115: 178-182.

Micheyl, C., Carbonnel, O., and Collet, L. (1995). Medial olivocochlear system and loudness adaptation: Differences between musicians and non-musicians. Brain and Cognition, Vol.29, 127-136.

Mountain, D.C. (1980). Changes in endolymphatic potential and crossed olivocochlear bundle stimulation alter cochlear mechanics. Science, Vol. 210, 71-72.

Sandman, D.E., Weiler, E.M., and Pederson, L.M. (1982). Perstimulatory adaptation by magnitude estimation: Effects of time of day, level of continuous tone, and relative level of periodic increments. Journal of Auditory Research, 22, 65-69.

Scharf, B. (1983). Loudness adaptation. In J. Tobias & E.D. Schubert (eds), Hearing Research and Theory (vol. 2, pp. 1-56). New York: Academic Press.

Tannen, R.S., Weiler, E.M., Warm, J.S. Dember, W.N, and Simon, J.O. (2001). Monaural loudness adaptation for middle-intensity middle-frequency signals: the importance of measurement technique. Journal of General Psychology, Vol. 128, 385-399.

Veuillet E., Collet L., Duclauz R. Effect of contralateral auditory stimulation on active cochlear micro-mechanical properties in human subjects: dependence on stimulus variables. J Neurophysiol 1991; 65, 3: 724-35.

Ward, W. D., (1973). Adaptation and Fatigue, In J. Jerger (Ed.), Modern developments in audiology (pp. 301-344). New York: Academic Press.

Weiler, E.M., Cross, K., Boudouris, S., Boyce, S., Kretschmer, L., Sandman, D.E., and Steger, J. The persistence of loudness in speech frequencies influence of ecological context. International Society for Psychphysics. Oct. 2001. p. 665-670.

Weiler, E.M., Dou, H., Tannen, R.S., Sandman, D.E., Dember, W.N., and Warm, J.S. Evoked otoacoustic emissions, suppression and comparison to loudness constancy, at speech frequencies. J. Acoustical Society of America. June, 1998. p. 3052.

Weiler, E.M., Sandman, D.E., and Pederson, L.M. (1981). Magnitude estimates of loudness adaptation at 60 dB SPL. British Journal of Audiology, 15, 201-204.

Weiler, E.M., Sandman, D.E., Janson-Pinto, J., Dange, A.J., Dember, W.N., Warm, J.S. and Nelson, T. (2000). Ipsilateral loudness adaptation over multiple intensity levels. Journal of General Psychology, Vol. 127, 365-371.

Weiler, E.M., Gold-Smith, L., Sandman, D.E., and Warm, J.S. (1992). Four triggering factors in loudness adaptation. Journal of General Psychology, 119, 325-334.

A PERCEPTUAL MAXWELL'S DEMON

Willy Wong
Edward S. Rogers Sr. Dept. of Electrical and Computer Engineering,
University of Toronto, Toronto, Ontario, Canada, M5S 3G4

Abstract

The concept of entropy has permeated physics since the 19th century. Entropy provides a medium of approach to understand physics on the basis of observation. Since psychophysics concerns itself with studying and quantifying the act of observation, it is natural to expect entropy to fulfill a role in unifying psychophysics to mainstream physics. This article reviews the concept of entropy in physics and psychophysics, and attempts to show their equivalence through an 'entropy-information cycle'.

Nobel laureate physicist Werner Heisenberg was once quoted as saying: "Thermodynamics leaves classical physics... for it speaks about situations of observation; it does not speak about the system as it is, but about the system in a certain state of being observed" (Buckley and Peat 1996). This rather insightful remark suggests that physics is somehow bound together to the observer studying it. The role of observers in physics has achieved increasing prominence in the past 100 years and can be traced at least as far back as Newton through the debate of absolute versus relative motion.

Following Heisenberg, the connection that concerns this paper best derives itself from the work of Ludwig Boltzmann. Boltzmann was an eminent Austrian physicist from the nineteenth century who advocated the microscopic approach to the study of matter and energy. It was Boltzmann who introduced entropy as a measure of molecular disorder. This concept was later extended to a theory of communication by Claude Shannon. Entropy has also found application in psychology and psychophysics.

In this paper, I wish to demonstrate that entropy is fundamentally concerned with human observation and measurement. Entropy provides a currency from which psychophysics can be understood in terms of physics concepts. The main thesis of this approach is that entropy in physics and entropy in psychophysics are one and the same. They form an integral part of a larger entropy-information cycle. This equivalence suggests that psychophysics can ultimately be explored through the realm of physics.

Entropy in physics

To comprehend the significance of entropy in physics, we require some basic concepts in statistical physics. I provide here a brief overview. We begin with the notion of energy. When energy provides a means to do work, it is referred to as 'free energy'. Free energy F, in thermodynamics, is related to the entropy S through the fundamental relation

$$F = U - TS, \tag{1}$$

where U is the total energy of the system and T is the temperature. Holding U and T constant, we see that an increase in entropy (or disorder) in the system results in a reduction of free energy or energy available to do work. As the disorder of a system increases, the ability to do work decreases.

In statistical physics, it is important to distinguish between equilibrium and non-equilibrium states. A system in equilibrium implies that it has reached steady-state and its properties are stationary or time-independent. An example of a non-equilibrium system would be a container of gas where the air molecules are aggregated in a small part of the container. Over time the molecules undergo diffusion, and eventually equilibrium is attained when the molecules are evenly distributed under a state of maximum disorder.

The calculation of entropy for a system in equilibrium proceeds as follows. A statistical system can be characterised in terms of its macroscopic and microscopic properties. As an example, temperature is a macroscopic variable that has a microscopic basis in terms of the kinetic energy of the individual molecules. Given a single macroscopic state (i.e. a *macrostate*), there are a large number of microscopic configurations or states (i.e. a *microstate*) compatible with the same macrostate. Entropy is then calculated by taking the logarithm of the number of compatible states. For example, consider a system constrained at a particular temperature T. This temperature specifies uniquely the total internal energy in a closed system. The energy is then distributed amongst the molecules in the system. Note that the energy does not have to be distributed evenly – in fact counting the total number of ways of distributing the energy is a combinatorial problem. An important point to remember is that the microstates are not directly observable and hence the entropy quantifies the uncertainty as to which microscopic state the system is in.

In non-equilibrium conditions, the entropy is calculated on the basis of the well-known Boltzmann H function

$$H = \int f \log f \, d\omega \qquad S = -k_B H \tag{2}$$

where S is physical entropy and k_B is Boltzmann's constant. A crowning achievement of 19th century theoretical physics was the demonstration that the H function obeys the inequality $dH/dt \leq 0$. This in turn implies that in a closed-system the physical entropy increases or stays the same This was a significant advance for Boltzmann's time as Newtonian mechanics – the basis of physics in the nineteenth century – was known to be invariant to time-reversal. Invariance to time-reversal implies that all mechanical motion can be reversed without appearing unnatural. Since it was believed that the individual molecules themselves follow Newton's laws, it was a mystery why irreversibility appeared in statistical systems. Irreversibility or the second law of thermodynamics was referred to as the "arrow of time" and explains why the air in the room does not spontaneously converge to one side of the room.

More recently these ideas were confirmed via computer simulations involving dilute gases of hard disks (Orban and Bellemans 1967). The hard disks themselves undergo reversible Newtonian dynamics but the entropy of the system H was shown to decrease over time as predicted by Boltzmann's law (see Fig. 1). The simulations also demonstrated that the increased entropy of the system corresponded to an increase in observer uncertainty of the system configuration. The disks were initiated at t = 0 with identical speeds and known initial positions. This corresponds to a state of *known* macroscopic configuration and *known* microscopic configuration. Over time, this system was allowed to evolve through interaction and collisions into a state of maximal uncertainty in terms of microscopic configuration. The increase in uncertainty has important ramifications for perceptual entropy to be discussed later.

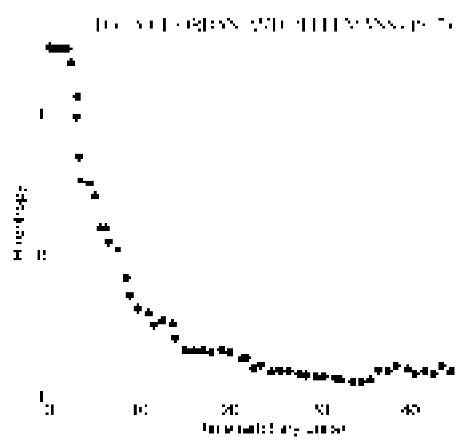


Fig. 1: Entropy of a dilute gas of hard disks

Entropy in psychophysics

The application of entropy or information to psychophysics is relatively well-known because of the work that has appeared on the magical number 7. However, we will not be utilizing this particular *extrinsic* use of entropy in psychophysics. Instead, we will focus on the approach pioneered by Ken Norwich (e.g. 1977, 1993).

Central to Norwich's theory is the association of uncertainty with perception. Perception is the act of reducing uncertainty; with certainty, perception ceases. The seminal equation that relates uncertainty to perception is given by

$$H = -\int p \log p \, dx \qquad F = kH \tag{3}$$

where F is the perceptual variable (like sensory magnitude, i.e. brightness), k is a constant and H the perceptual uncertainty. The utility of these equations in unifying classical psychophysics has been documented in a number of publications (e.g. Norwich 1993) and will not be covered here.

The resemblance of these equations to Boltzmann's equations was never intended although it provides the impetus to relate the two entropies later. But first we concentrate on deriving two crucial results from the perceptual H function. A closed form solution of the H function can be obtained under some very general conditions. Entropy is calculated on the basis of measurement uncertainty of a sensory signal. Uncertainty is due to the estimation errors that arise from limited samples of the signal. As with any measurement, increasing the number of samples increases the confidence in the estimate of the signal. As sample size increases, measurement error will approach a Gaussian pdf according to the central limit theorem and hence the entropy can be written as

$$H = \frac{1}{2}\log\left(1 + \frac{SNR}{m}\right) \tag{4}$$

where SNR is the signal to noise ratio and m represents the sample size. Note that this equation is identical to the well-known expression for channel capacity of a symmetric channel with Gaussian noise as derived originally by Shannon. m increases monotonically ($dm/dt \geq 0$) with the accumulation of samples over time. As m increases, so does the accuracy of the estimate and hence the uncertainty H (and F) decreases monotonically. Adaptation can therefore be interpreted as a gradual reduction of measurement uncertainty over time.

The second result is derived from the first. Taking the time derivative and holding SNR constant, we see that

$$\frac{dH}{dt} = -\frac{1}{2}\frac{1}{m^2 + SNR \cdot m}\frac{dm}{dt} \leq 0 \tag{5}$$

where the inequality is a consequence of $m \geq 0$ and $dm/dt \geq 0$.

This remarkable result shows a further parallel with physical entropy. A perceptual 'arrow of time' can be readily observed in any adaptation experiment at either a neural or a psychophysical level. The early moments aside, the adaptation curve falls monotonically with dF/dt and dH/dt both ≤ 0.

Marrying physical and psychophysical entropy

The aim of this paper is to show that physical and psychophysical entropy are in fact the same quantity. Before we proceed to this proof it is important understand why this should be expected. Indeed two systems can share a common mathematical description without being identical systems (c.f. diffusion phenomena and quantum mechanics). The answer to this question lies with a century old paradox in physics.

Scottish physicist James Clerk Maxwell illustrated deficiencies in the microscopic interpretation of the second law of thermodynamics with the following example (e.g. Leff and Rex 1990). Imagine a chamber containing two types of gas molecules, hot (fast) molecules and cold (slow) molecules. With reference to Fig. 2, the gas is initially well-mixed so that the system is in a high state of disorder (a). An intelligent demon (referred to as Maxwell's demon) installs a partition in the middle of the chamber (b). In the middle of the partition is a door which is controlled by the demon. The demon sits near the door and observes the gas molecules. When a fast traveling molecule from the right side moves towards the door, the demon opens the door to let the molecule pass. Conversely if a slow traveling molecule is moving left to right, the demon again opens to door to let the molecule pass. The door remains closed for all other molecules. That way, the original gas mixture can over time be separated into two phases (c). This is a violation of the second law. A consequence of this process is that useful work can now be extracted from the system (i.e. S is decreased and hence free energy F becomes available).

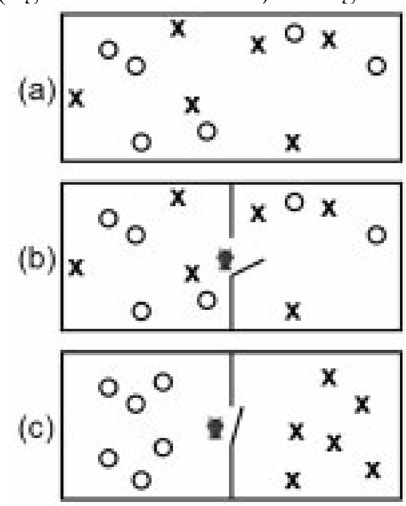


Fig. 2: The Maxwell's demon problem

This apparent violation of the second law has spawned a number of papers seeking to resolve this paradox. Perhaps the most compelling (and famous) explanation is the one provided by Leo Szilard. Szilard in his landmark paper argued that the analysis of Maxwell's demon is incomplete as it ignores the demon in the analysis (Szilard 1929). A real-life demon cannot operate without some means of measuring molecular speeds and the energy cost to perform this discrimination is equal to the free energy made available from separating the two molecule types. This cost can be calculated to be $k_B T\, ln(2)$ per molecule. Correspondingly, the information required by the demon to operate is $k_B\, ln(2)$ per molecule which is precisely the amount of entropy lost by the system. The analysis by Szilard highlights the connection between entropy and the acquisition of information.

Brillouin later termed this gain in information/loss of entropy as "negentropy" or negative entropy (Brillouin 1962). By including negentropy in the analysis, balance can be achieved when the cost of the demon is included. There have been suggestions that many real life problems (e.g. cellular ion channels, chemical bonding) follow a similar analysis in terms of information and energy exchange. Von Neumann postulated that computers calculating a "bit" must incur a minimal energy cost of $k_B T\, ln(2)$ joules, although real computing machines have been found to dissipate far more energy. It is clear that analysing the Maxwell's demon problem for even simple systems poses some serious difficulties.

The main difficulty in extending the Maxwell's demon problem to intelligent beings has been the lack of a clear connection between information and sensory function. With the developments of the entropy theory in psychophysics described earlier, one part of the puzzle is now complete. Indeed there has been work preceding this paper which explores the equivalence of perceptual entropy and physical entropy under equilibrium conditions. These studies show (a) that under some conditions both entropies are equal up to a constant additive term (Wong, 1993) and (b) that physical constants can be extracted directly from psychophysical data (Norwich 2001).

In the next section I outline a thought experiment which attempts to show the equivalence of perceptual and physical entropy in a general non-equilibrium case through an entropy-information cycle.

A thought experiment

The gist of the argument is as follows. A well-controlled signal is transmitted to a recipient who then decodes the signal on the basis of its characteristics. However the signal degrades as consequence of the second law prior to being received. The information required to decode the message or signal is equal to the degradation or increase in entropy of the signal. That is, the perceptual information or entropy required is equal in magnitude to the entropy change in signal due to the second law.

The thought or *gedanken* experiment involves an experimenter who creates a signal from a dilute gas. The gas is prepared in a precise manner so that the initial speeds of all the molecules are identical and the molecules are located in a small configurational volume (i.e. a system with low entropy). Before this signal is transmitted (i.e. allowed to be observed), it evolves according to the second law. In other words, the molecules diffuse outward to fill the gas container and the velocities change as a result of collisions. Each molecule undergoes a large number of identical collisions and the entropy of the gas increases. In particular, the velocity distribution approaches a normal distribution asymptotically via the central limit theorem. This increase of entropy can be used to extract work from the system (if a piston is added) but the changes of the system itself result in a loss of information: the experimenter can no longer keep track of the microscopic configuration of the system. That is, information about the macrostate is retained but knowledge of the microstate is lost. The information change is equal to the difference between the initial and the final entropy when the system reaches equilibrium.

Next the system is observed by the recipient. The message is contained in the macrostate of the signal but since the macrostate is unknown to the observer of the system, the observer must sample the microscopic configuration of the signal to obtain this information. The uncertainty or acquired information is calculated according to the method described earlier. Sampling entails that the certainty in the estimate of the macrostate will approach a normal distribution asymptotically according to the central limit theorem. That is, the process of information acquisition mirrors the process of information loss due to the second law. Perceptual information is equivalent to physical entropy.

Discussion and conclusion

The entropy-information cycle described above is analogous to the cycle proposed by Szilard, with the notable difference that the mechanism of information gain is now understood in terms of a sensorial basis. The main result which underlies this paper is the equivalence of perceptual and physical entropy. The thought experiment presented above seeks to argue that the Boltzmann irreversibility graph (c.f. the graph showing the entropy of hard disks) can in fact be mapped one-to-one to an adaptation curve. The only difference between a physical entropy curve and an adaptation curve is attributable to differences in offset, scaling and time scale.

The acquisition of perceptual information naturally requires the consumption of energy. According to Szilard's analysis, the minimal energy requirement to acquire information should be comparable to the energy made available to the system. This is a difficult matter to analyse in psychophysical experiments as most experiments cannot be carried out in such high precision to tease out the small amounts of energy consumed. However it is useful to note that for almost all modalities, the absolute threshold has a minimal energy requirement via Bloch's law (power x exposure = energy constant). Moreover a similar analysis can be carried out at the neural level in vision where single photons are detected. At body temperature, the energy consumption to make a binary decision is $k_B T \ ln(2) \approx 10^{-21}$ J. The energy released from a yellow-green photon interacting with a photoreceptor is on the order of $E = h\upsilon \approx 10^{-19}$ J. Thus the energy derived from the light signal is more than a hundred fold what is required for the neuron to discriminate between the presence or absence of a light signal. Given that a single photon provides more than sufficient energy for the discrimination task, why is the neuron only able to respond to single photon events statistically? The answer may lie with the quantum efficiency of the eye. Since not all photons that enter the eye reach the retina, the percentage that actually initiate a sensory event are on the order of 1% (van Meeteren 1978) which may account for why single photons can only be detected statistically.

Acknowledgements

The author would like to acknowledge many years of discussion on this topic with Professor Kenneth Norwich.

References

Brillouin L. (1962). *Science and information*. New York: Academic Press.

Buckley P. and Peat F.D. (1996). *Glimpsing reality*. Toronto, Canada: University of Toronto Press.

Leff H.S. and Rex A.F. (Eds.) (1990). *Maxwell's demon. Entropy, Information, Computing*. Bristol : Adam Hilger.

Norwich K.H. (1977). On the information received by sensory receptors. *Bulletin of Mathematical Biology, 39,* 453-461.

Norwich K.H. (1993). *Information, sensation and perception.* New York: Academic Press.

Orban J. and Bellemans A. (1967). Velocity-inversion and irreversibility in a dilute gas of hard disks.

Physics Letters A, 24, 620-621.

Szilard L. (1929). On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. Zeitschrift fuer Physik, 53, 840-856. English translation in Leff H.S. and Rex A.F. (Eds.) *Maxwell's Demon. Entropy, Information, Computing.* Bristol: Adam Hilger.

van Meeteren A. (1978). On the detective quantum efficiency of the human eye. *Vision Research, 18,* 257-267.

TIMING IS A NATURAL SECONDARY TASK: AN EXPLANATION OF THE ASYMMETRICAL INTERFERENCE IN TIMING

Dan Zakay
Department of Psychology, Tel-Aviv University, Israel.
dzakay@post.tau.ac.il.

Abstract

When timing is conducted simultaneously with a concurrent non temporal task, in most cases its performance is harmed while the performance of the other task is not influenced. This effect is termed the asymmetrical interference. In the present study it is argued that this phenomenon is caused by the way the executive resource allocation function treats timing under normal conditions. The argument is that whereas timing is naturally treated as a secondary task the concurrent non temporal task benefits from being treated as a primary task. Empirical evidence supports the argument by showing that the asymmetrical interference effect can be eliminated or even reversed in favor of timing if participants are instructed to treat timing as a primary task.

Brown (1997) presented the asymmetrical interference effect in timing. Based on an extensive literature review, it was shown that in most cases when prospective timing is performed simultaneously with a non temporal information-processing task, timing suffers more than the other one in terms of accuracy and of variability. The exceptions are non temporal tasks which include components of mental arithmetic. Brown's (1997) conclusion was that the resources used for timing are the same resources utilized by the executive function of working memory, as modeled by Baddely (1992). In the present study, a different approach is taken, an approach which does not necessarily contradicts the working memory resources explanation, on the one hand, but is not dependent on it on the other. In order to understand the new explanation a general model of prospective timing will be first presented. In the last part empirical evidence supporting the suggested explanation will be described.

Prospective timing takes place when one allocates attentional resources to timing in real time, while some target interval is actually experienced. It can be claimed that prospective timing is a dual –task condition because attention must be shared between temporal and non temporal information processing and attending to time requires access to some of the same resources that non temporal tasks use. For this reason models of prospective duration judgment are based on the assumption that prospective duration is a function of the amount of resources directly allocated for temporal information processing (Block & Zakay, 1996; Macar, Grondin & Casini, 1994, Zakay, 1999; Zakay &Block, 1996). These models often assume that signals which reflect the passage of time are accumulated in a cognitive counter. If a person focuses more attention on temporal information processing, more time signals are processed. If a concurrent nontemporal task is more demanding, fewer attentional resources are available for temporal information processing and fewer time signals are accumulated in the cognitive counter, whereby prospective duration judgment reflects the total number of accumulated signals. A direct result of this process is that the higher the difficulty of a concurrent non temporal task, the shorter experienced duration is. This conclusion is supported by several experiments (Zakay and Block,1997; Zakay and Fallach, 1984).The dependency of prospective duration judgment on the allocation of attentional resources, makes it vulnerable to attentional distractions. (Zakay,1992). The asymmetrical Interference effect (Brown, 1997) seems to be another result of the dual-task nature of prospective duration judgment.

However, the puzzling question, which remains to be explained, is why timing is, in most cases, the task which suffers from the competition over resources.

Timing is a natural secondary task.

We assume that under regular condition, the executive function which is responsible for determining a resource allocation policy has a natural tendency to prefer non temporal tasks over temporal ones and to allocate the first ones the needed amount of resources as required for optimal performance. In other words, it is argued that whenever timing does not have an important function in a specific context. It will be treated as a secondary task. The main reason for this phenomenon is rooted in the unique type of relationship between timing accuracy and the amount of attentional resources allocated for timing. Whereas for most non temporal tasks the more resources allocated for its performance, the better performance is, in timing this is not the case. Allocating too much resource might result in overestimation of target durations. Because of this there is no advantage in investing too many resources in timing. Due to this phenomenon, it is more optimal for the cognitive system to prefer non temporal tasks over temporal ones while determining the allocation policy and this explains the asymmetrical interference effect. However, when timing has a significant meaning the allocation policy is biased towards allocating more resources to it, even though this will not contribute to timing accuracy. In such cases the direction of the interference effect should be reversed. Two real life examples, waiting and time pressure, can illustrate the argument. While waiting for an important meeting, the experience is of a subjective time which runs much faster than does objective time. It is very difficult under this condition to be able to perform effectively any non temporal task. A similar condition is that of time stress like when one has a strict deadline to meet. It is well known that performance under time stress conditions deteriorates, partly due to shortage of resources because most of them are allocated for timing (Zakay,1993).

A careful examination of the method used by Brown (1997) reveals that he ordered participants to treat timing and concurrent non temporal tasks as equal in importance. A literature survey further reveals that usually either timing or non temporal tasks are treated as equal or that non temporal tasks are, implicitly or inexplicitly, treated as primary tasks. Under both conditions timing will be treated as a secondary task and an asymmetrical interference effect is expected to be found.

We predict that if participants will be instructed to treat timing as a primary task and a concurrent non temporal task as a secondary task, the asymmetrical interference effect will be eliminated or reversed in its direction. Zakay and Bibi (2002) conducted an experiment to test this prediction.

This experiment is summarized in the next section.

Method

Forty six first year social sciences' students participated in the experiment in partial fulfillment of course requirement. All participants were right handed with normal or corrected to normal vision. None of the participants had any prior knowledge about cognitive processes of timing.

The experiments were computerized and conducted on an IBM compatible PC, with a 17" super Vega graphic screen, a clock card, a keyboard and a mouse. All stimuli were presented on the computer screen and all participants' responses were performed by using either the keyboard or the mouse. Software programmed for the experiment measured, timed and recorded all stimuli and responses.

Experimental tasks

The timing (TT) task was performed by pressing a pre- determined key (the letter A) either every 2 sec. (TT2) or every 5 sec. (TT5). Each key press marked the termination of an interval and the beginning of a new one. The task was performed as a single task and was demarked by two tones, one in the "start" and one in the "end" points.

The tracking task (TR) was identical in its basic features to the tracking task used by Brown (1997, exp.1) and was chosen because a significant asymmetrical interference effect was found with it. However, whereas Brown utilized a special device, in the present study a PC was used. This required some changes in the speed of the target stimulus. In the present study two types of tracking tasks were used, an easy (TRE) and a difficult (TRD) tracking tasks. The target completed a whole cycle across a triangular track with 13.33 or 7.40 sec. in the TRE and TRD tasks, respectively. These cycles' time did not enable participants to use it as clues for producing the required intervals in the timing tasks. Tracking was done by moving the cursor which was controlled by the mouse. Participants were required to keep the cursor on the target (a black circle) as much time as possible. "Time on target" was used as a measure for tracking accuracy.
Each tracking task (TRE and TRD) were performed as single tasks.

Four combinations of dual tasks could be created by combining TT2, TT5, TRE, and, TRD. In dual tasks conditions both timing and a tracking task had to be performed simultaneously.

Three priority conditions were used:
Timing as a primary task and tracking as a secondary task.
Timing as a secondary task and tracking as a primary task.
Equal priority for timing and for tracking
Following Zakay (1992) priority was manipulated by giving participants different instructions regarding how they should treat each task.
Overall, 12 experimental conditions were obtained by presenting participants with the four dual tasks, each with three priority instructions.

The classical single-dual task paradigm was employed.
Each participant was tested individually and performed all 12 tasks in 3 different sessions. Single tasks were performed in a random order in the first session in order to prevent a potential benefit of training on its performance. The dual tasks were performed in a random order in the other two sessions.

The main dependent measures which were computed for each participant were: Timing accuracy, variability of time production, time on target and variability of time on target. The dependent measures were computed separately for each one of the 12 experimental conditions.

Summary of the main findings

Overall the results supported the tested hypothesis. A comparison between single tasks and respective dual tasks' performance revealed the existence of asymmetrical interference effect in favor of the tracking task when participants were instructed to treat tracking as a primary task or when timing and tracking were treated as equal in importance. However, when participants were instructed to treat timing as the primary task asymmetrical interference effects were not found and in some cases a reversed effect in favor of timing was obtained in timing variability.

Discussion

The findings obtained support an explanation of the asymmetrical interference effect based on the assumption that unless timing has a significant meaning or role it is naturally treated by executive control function in dual task conditions as a secondary task. Further support for this explanation was obtained by Block, Zakay and Richmond (2003), who found that when coupled with an automatic nontemporal task, timing performance was unharmed in the dual task condition compared to single task condition.

The findings also support attentional models of prospective timing and indicate the importance of paying attention to the instructions given to participants in timing experiments in terms of its potential impact on resource allocation policy.

.

Acknowledgement

This research was sponsored by a grant from the Basic Research Fund of the Israel Academy of Science.

References

Baddeley, A.D.(1992). Working Memory. *Science*, 255, 556-559.

Block, R.A., and Zakay, D. (1996). Models of Psychological Time Revisited. In Helfrich, H.(Ed.), *Time and Mind*. Hogrefe and Hubar, Kirkland, WA. pp. 171-195.

Block, R.A., Zakay, D., and Richmond, A.S. (2003, November). Automatic and Intentional processes in Face Encoding. *Poster presented at the meeting of the Psychonomic Society*, Vancouver, Canada.

Brown, S.W. (1997). Attentional Resources in Timing: Interference Effect in Concurrent Temporal and Nontemporal Working Memory Task. *Perception & Psychophysics*, 59, 1118-1140.

Macar, F., Grondin, S., and Casini, L. (1994). Controlled Attention Sharing Influences Time Estimation. *Memory and Cognition*, 22(6), 673-686.

Zakay, D. (1992). The Role of Attention in Children's Time perception. *Journal of Experimental Child Psychology*, 54, 355-371.

Zakay, D. (1993). The Impact of Time perception Processes on Decision under Time Stress. In: O. Svenson, and A.J. Maule, (Eds.) *Time Pressure and Stress in Human Judgment and Decision Making*. New York: Plenum press.

Zakay, D. (1999). Gating or Switching? Gating Is a Better Model of prospective Timing. *Behavioral Processes*, 50, 1-7.

Zakay, D. and Bibi, A. (2002). *Temporal Judgment, selective Attention and Attention Allocation Policy*. Ph.d Dissertation. Department of psychology. Tel-Aviv University.

Zakay, D. and Block, R.A. (1996). The Role Of Attention In Time Estimation processes. In: M.A. Pastor, and J. Artieda, (Eds.). *Time, Internal Clocks and Movement*, pp. 143-164.

Zakay, D. and Block, R.A. (1997). Temporal Cognition. *Current Direction in Psychological Science*. 6. 12-16.

Zakay, D. and Fallach, E. (1984). Immediate And Remote Time Estimation: A Comparison. *Acta Psychologica*, 57, 69-81.