Proceedings of the 36th Annual Meeting of the

International Society for Psychophysics

Fechner Day 2020

{Online}

October 19-22, 2020

Published by the International Society for Psychophysics, 2020.

Cover Image: Markus Spiske

Cover design: Jordan Richard Schoenherr

Title: *Proceedings of the 36th Annual Meeting of the International Society for Psychophysics Fechner Day 2020*

Edited by Jordan Richard Schoenherr, Timothy Hubbard, William Stine, and Craig Leth-Steensen for the International Society for Psychophysics.

01010100 01100001 01100010 01101100 01100101 0100000 01101111 01100110 0100000

Table of Contents

01100011 01101111 01101110 01110100 01100101 01101110 01110100 01110011

Preface

Welcome to the 36th Annual Conference of the International Society for Psychophysics, Fechner Day 2020!

As a member of the Executive Committee, I am pleased to welcome you to Fechner Day 2020. Normally, we meet every year across the globe, catch-up over food and drink, and discuss research. Fechner Day represent a unique experience where old friends can meet, debate the theories and methods that consume our daily lives, and engage with new members who bring new insights and perspectives to our research.

As you will recall, it was originally our intention to hold Fechner Day in Ottawa, Canada. The organization consisting of Caroline Blais, Charles Collin, Craig Leth-Steensen and myself were ready to welcome you to the Lord Elgin in downtown Ottawa. Unfortunately, due to the global pandemic, the Executive Committee and myself thought it best to move the conference to an online venue, in order to reassure membership that Fechner Day would continue in some form and that we would be able to safeguard the health of our membership.

Of course, as the primary organizer, I have ***many*** people to thank. First and foremost, Natalia Postnova, who has helped me organize this online conference as ISP's Scientific Communications Officer. Her efforts, ideas, and insight have been greatly appreciated throughout this … unconventional process. Of course, I would also like to thank the Organizing Committee from the Ottawa conference and others such as Tim Hubbard and Bill Stine for forming the Program Committee for this online conference. As was evidence from many conversations I had leading up to the conference, the pandemic placed considerable stress on our academic community as we adapted to the new demands of online teaching. Their time and efforts were greatly appreciated.

Of course, I also need to thank the omnipresent, often silent partner of the Executive, Zhuanghua Shi, as he was instrumental in setting up the webpage and posting on the ISP main page as well as Rosanna Tristao and Mark Elliot for early offers of assistance and suggestions.

Finally, as members of the Organizing Committee, I cannot forget to thank our President, Kazuo Ueda for his guidance during this time. As well as the efforts of Leah Fostick, and Wolfgang Ellermeier for this assistance with organizing and establishing the registration process, respectively.

I hope that the conference proves to be a sufficient stopgap until we meet again.

I wish you and your families health and happiness in 2020! See you in Ottawa in 2021!

Jordan Richard Schoenherr, PhD, Conference Chair

Department of Psychology / Institute for Data Science, Carleton University

Committee Membership

<u>Organizing / Executive Committee</u>

Kazuo Ueda (President)

Wolfgang Ellermeier

Leah Fostick

Natalia Postnova

Jordan Richard Schoenherr (FD2020 Chair)

<u>Program Committee</u>

Timothy Hubbard

Jordan Richard Schoenherr

Craig Leth-Steensen

William Stine

<u>Proceedings Committee</u>

Natalia Postnova

Jordan Richard Schoenherr, Chair

01010011 01100011 01101000 01100101

Schedule

01100100 01110101 01101100 01100101

<u>**Symposia Day 1 (Oct 19)**</u>

Looking Forward / Looking Back:

The History and Future of Psychophysics

Start Time: 08:00 EST (Americas):: 14:00 CEST (Europe/MidEast):: 21:00 JST (Japan)

Time (EST) **Start Stop**	**TITLE**	**AUTHORSHIP**
08:00 *08:02*	**Welcome Message**	Kazuo Ueda
08:02 *08:05*	**Session Introduction**	Jordan Richard Schoenherr
08:05 *08:15*	**Research Problems in the History of Psychophysics**	David Murray
08:15 *08:30*	**The Wave Theory of Difference and Similarity:** **New Perspectives**	Stephen W. Link
08:30 *08:45*	**The Future of Psychophysics: A Personal View**	Timothy L. Hubbard
08:45 *08:55*	**Interactive Break**	
08:55 *09:10*	**Black Boxes of the Mind:** **From Psychophysics and Explainable Artificial Intelligence**	Jordan Richard Schoenherr
09:10 *09:25*	**The Origin of Vierordt's Law**	Zhuanghua Shi and Stefan Glasauer
09:25 *09:40*	**Just Noticeable Differences in Neuroscience and Society:** **The Case of Emil du Bois-Reymond**	Gabriel Finkelstein
09:40 *09:45*	**Closing session**	

Symposia Day 2 (Oct 20)

Auditory Perception and Discrimination

Start Time: 07:00 EST (Americas):: 13:00 CEST (Europe/MidEast):: 20:00 JST (Japan)

Time (EST) **Start** **Stop**	**TITLE**	**AUTHORSHIP**
07:00 *07:05*	**Session Introduction**	Jordan Richard Schoenherr
07:05 *07:20*	**Open Science - Opportunities for Advancing Psychophysics**	Diana Kornbrot
07:20 *07:25*	**Extended Question Period**	
07:40 *07:55*	**Functional Differences in the Neural Substrates of Auditory Cognition as a Consequence of Music Training**	Naomi du Bois, José M. Sanchez Bornot, Dheeraj Rathee, Kong Fatt Wong-Lin, Girijesh Prasad, and Mark A. Elliott
07:55 *08:10*	**Can You Hear What I Feel?** **Simulating High-Frequency Hearing Loss** **Mimics Effects of Aging and Tinnitus in Emotional Speech Perception.**	Yehdua Dor, May Rosenblum, Dor Kenet, Vered Shakuf, Daniel Algom, and Boaz M. Ben-David
08:10 *08:15*	**Interactive Break**	
08:15 *08:30*	**Cross-modal commutativity of brightness and loudness productions**	Wolfgang Ellermeier, Florian Kattner, And Anika Raum
08:30 *08:45*	**Temporal Summation for Young and Older Adults**	Leah Foshtick
08:45 *08:50*	**Closing Session**	

Symposia 3 (Oct 21)

Perceptual Integration, Pain, and Temporal Perception

Start Time: 08:00 EST (Americas):: 14:00 CEST (Europe/MidEast):: 21:00 JST (Japan)

Time (EST) Start Stop	TITLE	AUTHORSHIP
08:00 08:05	**Session Introduction**	Jordan Richard Schoenherr
08:05 08:20	**Manipulating Self-Efficacy Beliefs in Basketball with Temporal Pressure**	Daniel Fortin-Guichard, Christian Jacques, Annie Goulet, Simon Grondin and Isabelle Giroux
08:35 08:50	**Neuroarcheology of pain sensation and perception: challenges to theoretical models**	Adamsom-Macedo and Rosana Tristão
08:50 09:05	**The Pythagorean Comma and the Stretched Octave: A Surprising Similarity**	Timothy L. Hubbard
09:15 09:25	**Interactive Break**	
09:25 09:40	**Motion Induced Blindness: Luminance Contrast Sensitivity to Increments and Decrements in the Presence of a Motion Mask**	Rebecca A. White, Dan Swift, Sofia C. Lombardo, Erika S. Wells, John E. Sparrow, Abby Oostendorp, Joseph A. LaBarre, Andrew J. Kitt, and Wm Wren Stine
09:40 09:55	**The Characteristic Properties of Cognitive Processes with Perceptually Integral Stimuli**	Yanjun Liu and James Townsend
10:10 10:25	**Who Gives a Damn about Theory: Taten sagen mehr als Worte**	Willy Wong
10:25 10:40	**Modelling Subjective Confidence with an Associative Memory Network**	Jordan Richard Schoenherr
10:40 10:45	**Closing Remarks**	

Symposia 4 (Oct 22): Business Meeting

Start Time: 07:00 EST (Americas):: 13:00 CEST (Europe/MidEast):: 20:00 JST (Japan)

-

Abstracts and Papers

:: Listed Alphabetically by First Author ::

Neuroarcheology of pain sensation and perception: challenges to theoretical models

Elvidina Nabuco Adamsom-Macedo[1] and Rosana Tristão[2]
[1]Emeritus Professor of Maternal-Infant Mental Health, formerly School of Health, University of
[2]Wolverhampton, England, United Kingdom
Faculty of Medicine and University Hospital, University of Brasilia, 70910-900, Brasilia, Brazil

Abstract— The aim of this paper is to reflect upon neonatal pain (neonatal preterm babies focused), the appearance and development of the field of neuroarcheology and its possible impact and implications to the field of Neonatal Pain theoretical models. Within the theoretical rationale of Neonatal (Health) Psychology - based on G. Gottlieb's theory of System Development (Gottlieb, 1991)- the author(s) will also reflect upon the role of touch on the alleviation of neonatal pain, how intensive care babies responded to it, neuro-archeology of pain and possible challenges to theoretical models of pain.

Keywords— Pain, neuro-archeology, touch, development

I. Introduction

Richard L. Gregory in 1986 coined the term Neuroarchaeology in an editorial of that title in the Journal Perception (see Laughlin, 2015, p 346, Note 1). In 25th March 2001 Gregory addressed the Swiss National Science Foundation Symposium 'Diseases of the Nervous System' with Communication entitled "Neuro-Archaeology: Some Speculations on Evolution". The author(s) of this Chapter will attempt to reflect upon Gregory's Speculations in relation to Pain and Touch as well as other authors engaged with neuroarcheology.

At his presentation in 2001 Gregory pointed out that Charles Darwin's theory of Evolution has cast light on every aspect of biology and, neurology is no exception. He stated that an essential feature o Evolution is its arrow through time: it cannot "go back to the drawing board" for new designs. He continues by saying that there are seldom entirely new structures; instead existing structures take on new and sometimes very different functions albeit modifications that follow may be quite slow. Do we live in the present with outdated maps from the past? Gregory asked. This indeed is fascinating as new research appears to be pointing at such direction!

Gunnar et colleagues launched a publication entitled "When damaged the adult brain repairs itself by going to the beginning" (Poplawski et al., 2020). These researchers at the University of California San Diego, School of Medicine, with colleagues elsewhere reported that when adult brain cells are injured, they revert to an embryonic state. In their newly adopted immature state, the cells become capable or re-growing new connections that under the right conditions can help to restore lost functions.

Gregory also referred to Arnold Gesell (Gesell; Amatruda, 1971)(first edition at 1945) and reflect on his 'Embryology of behaviour' (Introduction p. xiii) when he says "In the biological perspective, the newborn infant is an extreme ancient for he has already traversed most of the stages of his long, racial evolution." We can spend hours pondering about the reality and beauty of Gesell's statement. Continuing his reference to Gesell, Gregory highlighted his Chapter 5 – The Archaic Motor System, as a good starting point for seeing the "archaeological" time – layers of muscles and their functions. The oldest muscles are for posture (the basis of behaviour). Posture changed over hundreds of millions of years from horizontal to vertical. In order to accommodate new posture muscles and neural organizations changed; greatly revised strategies were needed so that moving around and performing skills was possible.

II. ONTOLOGY AND PHYLOGENY OF PAIN SENSATION AND PERCEPTION

The development of skills is seen in terms of innate ontology, as well as individualized learning. Thus, on page 52 (Gesell's chapter 5) reads "Complicated action patterns whose components were ontogenetically and physiologic developed over long reaches of time are telescoped into a single moment of behavior." Relating posture to behavior, Gesell cites the classical studies of motor development for swimming behaviour and responses to touch stimuli of salamanders by Coghill (Coghill, 1914). Coghill distinguished innate development from learning and also from maturation, which may require active behaviour though it is not learning.

The very term neuro-archaeology combines two powerful areas of human knowledge: neurology and archaeology. This cross-disciplinary field has boosted new scientific concepts born from the interaction from old human cognitive traces in archaeological findings. These have allowed both areas to draw the blueprint of the human mind tracing back millions of years. The approach proposes the construction of an "analytical bridge between brain and culture, by putting material culture, embodiment, time and long term change at centre stage in the study of mind" (Malafouris, 2010). The cognitive archaeology, as this area is named, proposes concepts as metaplasticity that encompass both cultural and neurophysiological evolution entwined in a complex process that is now been mapped. Concepts has evolved in this field and the area has changed titles from psychopaleontology, neuroconstrutivism and probabilistic epigenetics, among others (D'Souza; KarmiloffSmith, 2017; Fry, 2006; Rinaldi; Karmiloff-Smith, 2017).

III. Touch as a Complex Sense

When, it comes to pain, one of the four subsystems of Touch besides temperature, proprioception and pressure specific receptors, we face a challenge to decipher which one comes first in the embryo life. Pain is defined by the International Association for the Study of Pain (IASP) as 'an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in 'pain is always subjective and is always a psychological state' (Anand; Stevens; Mcgrath, 2007; Merskey; Bogduk; International

Association for The Study of Pain, 1994). Nevertheless, researchers still face the challenge to understand how newborns are able to perceive pain and if they are able to discriminate different sources of pain. Pain measurements in newborns involve physiological and behavioural assessment, although it is not always clear whether they are measuring pain or distress (Bellieni et al., 2007).

Moreover, as pain is also a vital signal, one can propose that once life was stablished the vital signals must have emerged together in an architectural relation, structurally meaning, suggesting pain as the earliest function to be activated at the proto mind in development. Pain sensation as a response to nociceptive stimulus, among all vital signals, employs an important role as it also has a defensive function. The nociceptive response in newborn infants has been formally studied since Charles Darwin when he and Phillip Prodger published, in 1872, their book "The Expression of the Emotions in Man and Animals", which was re-edited posteriorly (Darwin; Ekman, 1998). They argued that all humans, and even other animals, show emotion through remarkably similar behaviours and this can be observed early in life proving the evolutionary and phylogenetic link for human pain response. Actually, the anatomical and functional foetal development for touch and its receptor subsystems is traced back to the 7th gestational week with avoidance response for the lips being touched as described by Bremer et al and Gottlieb in Tobach et al (Bremner; Lewkowicz; Spence, 2012; Tobach et al., 1971). Although, the foetal reaction of avoidance is known since then, we cannot affirm that the early foetus is able to react differentially to nociceptive stimulation, but we can hypothesize that the defensive response may be one of the oldest human response to the environment and also that it can be the spark that lights the mind.

References

Anand, K. J. S., Stevens, B. J., & McGrath, P. J. (Eds.). (2007). Pain in neonates and infants (3rd ed). Elsevier.

Bellieni, C. V., Cordelli, D. M., Caliani, C., Palazzi, C., Franci, N., Perrone, S., Bagnoli, F., & Buonocore, G. (2007). Inter-observer reliability of two pain scales for newborns. Early Human Development, 83(8), 549–552. https://doi.org/10.1016/j.earlhumdev.2006.10.006

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (Eds.). (2012). Multisensory development (1st ed). Oxford University Press.

Coghill, G. E. (1914). Correlated anatomical and physiological studies of the growth of the nervous system of amphibia. The Journal of Comparative Neurology, 24(2), 161–232. https://doi.org/10.1002/cne.900240205

Darwin, C., & Ekman, P. (1998). The expression of the emotions in man and animals. Oxford University Press.

D'Souza, H., & Karmiloff-Smith, A. (2017). Neurodevelopmental disorders. Wiley Interdisciplinary Reviews. Cognitive Science, 8(1–2). https://doi.org/10.1002/wcs.1398

Fry, I. (2006). The origins of research into the origins of life. Endeavour, 30(1), 24–28. https://doi.org/10.1016/j.endeavour.2005.12.002

Gesell, A., & Amatruda, C. (Strunk). (1971). The embryology of behavior: The beginnings of the human mind. Greenwood Press.

Gottlieb, G. (1991). Experiential canalization of behavioral development: Theory. Developmental Psychology, 27(1), 4–13. https://doi.org/10.1037/0012-1649.27.1.4

Gregory, R. L. (1986). Neuro-Archeology. Perception, 15(2), 93–94. https://doi.org/10.1068/p150093

Laughlin, C. D. (2015). Neuroarchaeology. Time and Mind, 8(4), 335–349. https://doi.org/10.1080/1751696X.2015.1111563

Malafouris, L. (2010). Metaplasticity and the human becoming: Principles of neuroarchaeology. Journal of Anthropological Sciences = Rivista Di Antropologia: JASS, 88, 49–72.

Merskey, H., Bogduk, N., & International Association for the Study of Pain (Eds.). (1994). Classification of chronic pain: Descriptions of chronic pain syndromes and definitions of pain terms (2nd ed). IASP Press.

Poplawski, G. H. D., Kawaguchi, R., Van Niekerk, E., Lu, P., Mehta, N., Canete, P., Lie, R., Dragatsis, I., Meves, J. M., Zheng, B., Coppola, G., & Tuszynski, M. H. (2020). Injured adult neurons regress to an embryonic transcriptional growth state. Nature, 581(7806), 77–82. https://doi.org/10.1038/s41586-020-2200-5

Rinaldi, L., & Karmiloff-Smith, A. (2017). Intelligence as a Developing Function: A Neuroconstructivist Approach. Journal of Intelligence, 5(2), 18. https://doi.org/10.3390/jintelligence5020018

Tobach, E., Aronson, L. R., Aronson, E., Beach, F. A., Shaw, E. S., & History, A. M. of N. (1971). The Biopsychology of Development. Academic Press. https://books.google.com.br/books?id=r7lsAAAAMAAJ

Functional Differences in the Neural Substrates of Auditory Cognition as a Consequence of Music Training

Naomi du Bois[1], José M. Sanchez Bornot[1], Dheeraj Rathee[1], KongFatt Wong-Lin[1], Girijesh Prasad[1], and Mark A. Elliott[2]

[1]Intelligent Systems Research Centre, University of Ulster, Magee Campus, Derry, NI, UK
[2]School of Psychology, National University of Ireland, Galway, Ireland

***Abstract*—** A magnetoencephalographic (MEG) investigation of differences in the dynamics of the auditory cognition system dependant on music training, as opposed to no experience playing an instrument, was conducted using a customised auditory priming paradigm1,2. This paradigm employed stimulus entrainment to evoke an auditory gamma-band response (aGBR), i.e. an oscillatory response in the range 30-70 Hz that is phase locked to the stimulus. Within a trial sequence, participants respond to the harmonic relationship between the carrier frequencies of an entrainment stimulus and the subsequent target stimulus, which carries either a harmonic or an inharmonic (deviant) tone in comparison. Neuroscientific research has demonstrated that musical deviants (syntactically irregular chords) elicit event related potentials/fields (ERPs/Fs) with negative polarity; the early right anterior negativity (ERAN) and the right anterior temporal negativity (RATN) responses, with peak latencies of ~200 ms and 350 ms, respectively, post stimulus onset 3,4. The focus of the time-frequency analyses was on the effect of stimulus entrainment on these passively evoked responses to deviant stimuli, as a function of music training.

The findings suggest that entrainment completely disrupts the ERAN response. Furthermore, the source location of ERF difference activity (the difference in activation elicited by the inharmonic deviant compared to the harmonic) during the later RATN time-window varies depending on group and entrainment condition; located in the left hippocampus and parahippocampal area following 33 Hz entrainment, the right superior frontal cortex following 37 Hz entrainment, and the left inferior temporal lobe following 39 Hz entrainment. The 33 Hz and 39 Hz ERF differences were driven by the non-musician group, as they were not evident for the musician group when ERF differences were compared for each group separately. These findings are supported by the reaction time (RT) analysis; significantly faster RT responses to the inharmonic compared to harmonic targets, were found for the same condition and group combinations. However, when ERF differences for musicians and non-musicians were compared, following 33 Hz entrainment a different source location was found to be significant for musicians, in the left middle frontal gyrus. Notably, a phase-amplitude coupling analysis revealed an overall 7 Hz phase modulation of gamma amplitude across the full range of frequencies examined (28 Hz to 42 Hz) for non-musicians, and was found to be statistically significant for approximately 33% of this group. A 7 Hz theta modulation effect following a 33 Hz entrainment supports two assertions from previous research; 1) non-musicians have faster RT responses to inharmonic, and/or slower responses to harmonic, auditory stimuli following 33 Hz entrainment 1,2 and 2) findings using a functionally similar visual priming paradigm, reveal faster RT responses dependent on a phase interaction between entrainment frequencies and a slower endogenous theta rhythm, of 6.69 Hz 5. The overall conclusion is that while stimulus entrainment in the gamma-band range interferes with the usual pattern of cortical responses involved in tone discrimination, significant effects of gamma entrainment are found more frequently in the non-music brain. In contrast, musicians demonstrate a greater range of interactions with slower brain rhythms, indicative of greater top-down control, suggesting musicians auditory responses rely more on top-down processes and thus the effect of stimulus entrainment in the gamma-range is reduced.

***Keywords*—** Music cognition, Stimulus entrainment, Auditory gamma-band response (aGBR), Phase-amplitude coupling (PAC).

References

[1]Aksentijevic, A. et al. J. Exp. Psychol. Hum. Percept. Perform. 37, 1628–1642 (2011)
[2]Aksentijevic, A. et al. Music Percept. An Interdiscip. J. 31, 316–322 (2014)
[3]Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. 4, 540–545 (2001)
[4]Rohrmeier, M. A. et al. Int. J. Psychophysiol. 83, 164–175 (2012)
[5]Elliott, M. A. Front. Psychol. 5, 1–13 (2014)

Can You Hear What I Feel? Simulating High-Frequency Hearing Loss Mimics Effects of Aging and Tinnitus in Emotional Speech Perception

Yehdua Dor[1,2], *May Rosenblum*[2], *Dor Kenet*[2], *Vered Shakuf*[2], *Daniel Algom*[1] *and Boaz M. Ben-David*[2,3]
[1]Department of Psychology, Tel-Aviv University
[2]Baruch Ivcher School of Psychology, Interdisciplinary Center, Herzliya
[3]Department of Speech-Language Pathology, University of Toronto

Abstract— Emotions in speech are presented primarily in the semantic (words) and prosodic (tone) channels. While young, healthy adults are heavily biased towards emotional prosody, older adults and people with tinnitus rely less on emotional prosody and more on emotional semantics. We focused on one factor shared by these two groups- changes in primary auditory processing. In the current study, young, healthy adults listened to spoken sentences carrying emotional content in both prosody and semantics that were digitally degraded to simulate reduced sensitivity to high-pitch sounds, typical to older adults. Simulating of hearing loss reduced the intelligibility of emotions in both channels for young, healthy adults. In addition, young listeners relied less on emotional prosody and more on emotional semantics, mimicking the response pattern of older adults and people with tinnitus. These results are discussed under the sensory information degradation theory.

Keywords— Emotion, Speech perception, Hearing Loss, Older Adults, Spoken Emotions

IV. INTRODUCTION

Successful processing and understanding of emotional information form a fundamental part of human communication. In the auditory modality, emotions in speech can be conveyed by two channels, the semantics channel (content of the words) and the prosodic channel (tone of speech and suprasegmental features: rhythm, stress, intonation etc.; Pell, Abhishek, Monetta & Kotz, 2011). The emotional information presented in the two speech channels can be congruent or incongruent: For example, when hearing the sentence "I feel wonderful today" spoken with angry prosody one may interpret the emotional message as being happy, angry, or as some combination between the two emotions.

The interplay between the processing of emotions from spoken semantics and prosody was tested in various studies, with mixed results in different populations. When young and healthy adults in their 20's were tested ("classic" samples in science) a growing body of literature points to the dominance of emotional prosody processing (e.g. Ben-David, Multani, Shakuf, Rudzicz & van Lieshout, 2016; Filippi et al., 2017; Lin & Ding, 2020). In other words, most young, healthy adults judge the sentence demonstrated above as being very angry (based on the angry prosody) and not very happy (underestimating the happy semantics of the words). In addition, when asked to selectively attend to one speech channel, prosodic information lead to larger failures of selective attention to the semantics than vice versa. These larger failures to inhibit the prosodic channel further indicate the existence of prosodic dominance in young, healthy adults.

Albeit these results, research in recent years suggests that dominance of emotional prosody is not necessarily universal. Two notable exceptions are the (non-clinical) population of older adults and the (clinical) population of people with Tinnitus. Older adults (60 to 75 years old) with clinically normal hearing (for their age), showed a slight bias towards semantics, judging spoken sentences based on the semantics a tad more than on their prosody (Ben-David, Gal-Rosenblum, van Lieshout & Shakuf, 2019; Dupuis & Pichora-Fuller, 2010). In addition, Ben-David et al. (2019) found larger failures of selective attention for older adults compared to young adults, reflecting the common notion in the literature on an age-related reduction in inhibition efficiency (Hasher and Zacks, 1988). Similar findings were noted for people with tinnitus. Oron et al. (2019) found the prosodic bias to be greatly reduces for people with tinnitus as compared to age-matched typically developed adults. Notably, selective attention was not affected by tinnitus. This distinction between older adults and adults with tinnitus hints on the possibility of additional cognitive processes that impact processing of spoken emotions in older age.

In order to investigate the perceptual basis underlying these differences in the integration of emotional information from different speech channels, we focused on one factor shared by the two aforementioned populations: changes in primary auditory processing, reflected by decreased hearing sensitivity. Regarding people with tinnitus, the exact causes of the condition remain unclear, but it is assumed to be related to cochlear damage or to decreased neural output from the cochlea (Schaette & McAlpine, 2011). As for older adults, abundant literature describes various age-related changes in the auditory system (Gordon-Salant, Frisina, Fay, & Popper, 2010). Specifically, in both clinically normal- and pathological hearing older adults, changes in hearing sensitivity and elevated hearing thresholds tend to be larger at higher sound frequencies, with hearing degradation slope becoming steeper as pitch increases (Davis, 1995; Engdahl et al.,2005).

In the current paper, we test whether changes in processing of emotions in speech may be attributes, at least in part, to sensory changes. This hypothesis is based on the theoretical framework of sensory information degradation (Lindenberger & Baltes, 1994; Schneider & Pichora-Fuller, 2000) attributing differences in "higher" cognitive processing to "lower" changes in basic sensory input. One prediction derived from such a perspective is that experimental stimuli can be simulated to mimic various basic sensory changes. Presenting such stimuli to young, non-clinical participants would result in different processing and a response pattern similar to that of older- or clinical- participants. In the field of speech perception, for example, Fontan et al. (2017) used digitally-manipulated spoken words and sentences to simulate age-related hearing loss, resulting in decreased intelligibility and comprehension in young normal-hearing adults. Füllgrabe (2020) used similar digitally-manipulated spoken sentences in memory tests administrated to young normal-hearing students, resulting in lower memory scores (thus mimicking allegedly cognitive decline).

In the current study, we adopted a similar approach to test the sensory basis of the integration of emotional information from the channels of semantics and prosody. We focused on one common aspect of hearing loss, elevated hearing thresholds to high-frequency sounds. Spoken sentences carrying emotional content in both semantics and prosody were digitally processed to simulate the hearing sensitivity of older adults. We hypothesized that younger adults listening to these simulated sentences would adopt a response pattern similar to that of older adults or people with tinnitus, namely, decreased prosodic dominance and increased reliance on emotional semantic information.

V. Method

Participants. Seventy-seven healthy, young adults (53 women, 24 men) at ages 18-30 years (M age=25.4 years, SD=2.5) who volunteered to participate in the study. All participants were native Hebrew speakers, with self-reported normal hearing, vision, cognitive state, as well as with no severe emotional symptoms (as assessed by the Depression, Anxiety and Stress questionnaire; Lovibond & Lovibond, 1995). Thirteen additional participants were excluded from analysis as they did not meet the abovementioned criteria.

General Procedure. After providing informed consent, participants filled a series of online questionnaires regarding demographic variables, health, emotional state and language proficiency. After filling the questionnaires, participants who met inclusion criteria were sent a link to the online version of the Test for Rating of Emotions in Speech (iT-RES, Ben-David et al., in press) and preformed the test on their personal computer using headphones.

iT-RES Stimuli and Procedure. The iT-RES consists of spoken sentences with emotional content in both semantics and prosody, either congruently (e.g. "Get out of my room" spoken with angry prosody) or incongruently (e.g. "I won the lottery" spoken with sad prosody). In this version of the test, the emotions of anger, happiness and sadness were used, as well a neutral baseline condition. Listeners are asked to rate to what extent they agree the speaker is angry, happy or sad using a 6-point Likert scale. In one block listeners are asked to rate the sentence as a whole, without referring to any specific speech channel, thus testing the integration of emotional information between the channels. In other blocks listeners are asked to rate the sentence based only on its prosody or semantics while ignoring the other channel, thus testing their ability to selectively attend to one channel alone. Additional information on the T-RES, as well as validation data for the online version of the test, can be found elsewhere (Ben-David et al., 2016; Ben-David et al., in press).

Participants were randomly assigned to one of two groups: in the control group (N=39), participants completed the iT-RES listening to the original recorded sentences of the test, reflecting normal hearing and good intelligibility of auditory stimuli. In the simulated hearing loss (SHL) group (N=38) the spoken sentences of the i-TRES were digitally degraded to reflect reduced sensitivity to high-frequency sounds. The level of degradation was determined based on a large population study assessing hearing thresholds in the adult population (Engdahl et al., 2005). Using the point of 1000Hz sounds as a baseline, intensity levels of sounds were decreased to simulate the hearing-loss gradient of older adults at the ages of 60-79 (with no self-reported hearing problems), thus mimicking reduced sensitivity to high-pitched sounds typical in older age. The final slope ranged from no intensity change at 500-1000hz up to a degradation of 40dB at 8000 Hz (see figure 1).

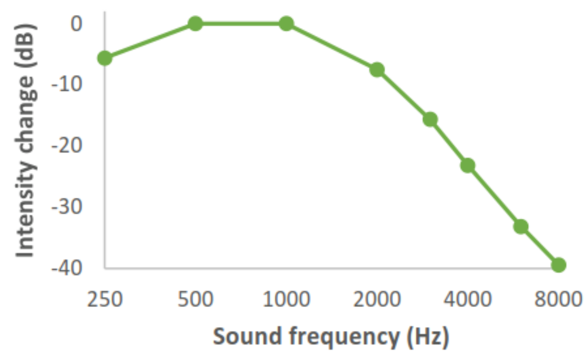

Figure 1. Hearing loss simulation filter.

Note. Digital filter simulating a gradient of hearing loss at high frequencies typical of ages 60-79 (based on Engdahl et al., 2005), using 1000Hz as a baseline.

VI. Results

Identification of Emotions. Identification of emotions was estimated by analyzing ratings of baseline sentences, in which one speech channel (e.g., semantics) carries neutral information, and the other (prosody) carries emotional information. A mixed-model repeated measures ANOVA was conducted, with emotion rating as the dependent variable, presence of target emotion in the sentence (target emotion present vs. absent), speech channel (prosody vs. semantics) and type of emotion (angry, happy or sad) as within-subject variables, and experiment group (SHL vs. control) as between-subject independent variable.

Results show a strong main effect for target emotion presence, $F(1,75)=1508$, $p<.001$, $\eta_p^2=.953$, indicating a very good ability to identify emotions in baseline sentences. A significant interaction between experimental group and emotion presence $F(1,75)=4.7$, $p=.033$, $\eta_p^2=.059$, indicated that emotion identification was slightly better in the control group than the SHL group. Notably, none of the interactions with the speech channel factor were found to be significant, indicating that the simulation of hearing loss affected the identification of emotions in both speech channels similarly.

Failure of Selective Attention. Selective attention was gauged by examining listeners' inhibition of information presented in a to-be-ignored channel. Again, a mixed-model repeated-measures ANOVA was performed with emotion rating as the dependent variable, speech channel (prosody vs. semantics), type of emotion (angry, happy or sad) and selective attention factor (the presence of target emotion in the to-be-ignored channel) as within-subject variables, and experimental group (SHL vs. control) as between-subject independent variable.

Results show significant failures of selective attention $F(1,75)=65.39$, $p<.001$, $\eta_p^2=.466$, that interacted significantly with the speech channel $F(1,75)=40.85$, $p<.001$, $\eta_p^2=.353$. These results indicate that failures to inhibit the prosodic channel were larger than failures to inhibit the semantics one. Notably, the experimental group factor did not interact with any of the other factors, suggesting that simulating hearing loss did not affect selective attention.

Integration of Emotions. The integration of prosody and semantics was gauged by analyzing performance on the general-rating task, in which listeners were instructed to judge the sentence as a whole, without a direction to one of the two channels. A mixed-model repeated-measure ANOVA was conducted with emotion rating as the dependent variable and experiment group (SHL vs. control) as between-subject independent variable. The within-subject independent variables were type of emotion (angry, happy or sad) and the presence of target emotion in any of the channels, using a linear-like factor: target emotion appears in both channels, only in prosody, only in semantics, or not at all.

The linear trend was found to be significant $F(1,75)=1743$, $p<.001$, $\eta_p^2=.959$, and to significantly interact with the experimental group $F(1,75)=4.46$, $p=.038$, $\eta_p^2=.056$ (see figure 2). A post-hoc analysis shows that the source for the interaction comes from the ratings of emotion-only-in-prosody and emotion-only-in-semantics sentences, $F(1,75)=5.57$, $p=.021$, $\eta_p^2=.069$. Namely, the prosodic bias, the advantage for rating prosodic information over semantic one, was larger for the control group than for SHL group (2.19 vs. 1.5 respectively).

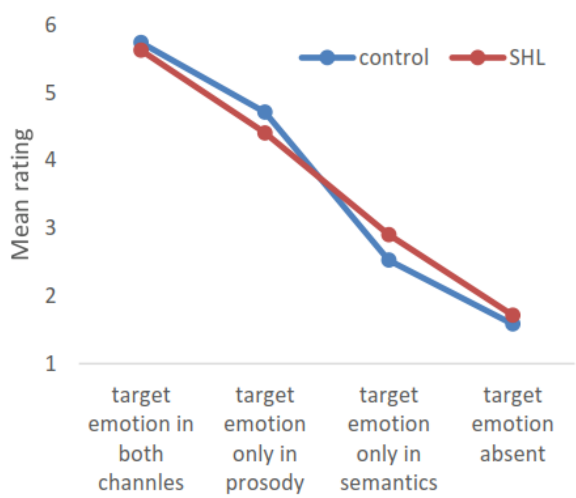

Figure 2. Integration of emotions from speech channels.

Note. Mean rating of emotions on a 1-6 Likert scale, when listeners are instructed to rate the sentence as a whole (general rating task).

VII. Discussion

The literature shows that young, healthy adults are biased to process emotions in speech based predominantly on prosodic cues; whereas older adults and people with tinnitus -- populations with deficits in primary auditory processing -- are at least less biased to process the prosody over the semantics of the sentence. In the current study we tested whether simulating reduced sensitivity to high-pitched sounds to young healthy adults (by digitally degrading spoken sentences) could mimic this trend. Namely, testing a sensory source for the aforementioned group-differences in channel integration.

Most importantly, simulated hearing loss (SHL) was found to reduce the prosodic bias, even for young healthy adults. SHL was also found to slightly reduce the intelligibility of emotions. Notably, as SHL affected the identification of emotions in both prosodic and semantic channels similarly, it appears that reduced prosodic bias for older adults and people with tinnitus does not arise from selective sensory degradation of prosodic information.

Our data can be seen as reflecting the theoretical framework of sensory information degradation (Lindenberger & Baltes, 1994; Schneider & Pichora-Fuller, 2000). It is feasible to assume that under adverse sensory conditions individuals tend to adopt a wider processing strategy, combining degraded information from several sources in order to perceive the full picture clearly. When basic auditory processing is impaired, one may be more willing to integrate emotions from both (degraded) prosody and (degraded) semantics, resulting in less prosodic dominance and more balance between the channels.

When asked to judge the sentence as a whole, listeners in the SHL group adopted a response pattern similar to that of older adults and people with tinnitus. Nevertheless, no difference was found between the groups regarding the effects of selective attention (similarly to people with tinnitus but unlike older adults). This points to the possible involvement of additional cognitive processes in emotion perception among older adults, such as inhibition deficits (Hasher & Zacks, 1988). Future research can test the effects of such cognitive processes on the perception of emotion in spoken prosody and semantics.

Caveats. As the study was conducted during COVID-19 social restrictions, participants were asked to perform the task in their home on their own personal computer using a dedicated tele-assessment tool, iT-RES. Participants were instructed not to change the loudness of the speakers during the test. Although we cannot be certain that all participants followed that instruction, it is important to note that even if some the listeners in the SHL group raised the volume to better hear the sentences, the effect of the hearing-loss gradient of reduced intensity to high-frequency sounds would still impact performance.

In the current study, we manipulated only one factor related to hearing loss -- reduced sensitivity to high-pitch sounds. This single change was sufficient to significantly affect the integration of emotions from speech channels. It is important to note that hearing loss related to aging or to clinical conditions influences several other auditory domains, which can be manipulated separately or jointly (e.g. Füllgrabe, 2020). Further research calls for the examination of these factors in relation to emotion perception in speech.

Conclusions. Simulating hearing loss was sufficient to impact both the identification and the integration of emotions from the prosodic and semantic channels, in young healthy listeners. This provides initial support for the sensory information degradation hypothesis. Namely, reduced prosodic bias might be related to reduced sensory input. Future studies should attempt to digitally impair spoken sentences to a larger extent – in an attempt to better mimic the impact of aging on the integration of channels. One may also note that selective attention was not affected by our sensory manipulation, suggesting that it may be represent a separate cognitive factor in aging (see Ben-David & Icht, 2017, 2018; Ben-David, Eidels & Donkin, 2014).

References

Ben-David, B. M., Eidels, A., & Donkin, C. (2014). Effects of aging and distractors on detection of redundant visual targets and capacity: Do older adults integrate visual targets differently than younger adults? *PLoS ONE* 9(12): 1-29. doi:10.1371/journal.pone.0113551

Ben-David, B. M., Gal-Rosenblum, S., van Lieshout, P. H., & Shakuf, V. (2019). Age-Related Differences in the Perception of Emotion in Spoken Language: The Relative Roles of Prosody and Semantics. *Journal of Speech, Language, and Hearing Research*, 62(4S), 1188-1202.

Ben-David, B. M., Multani, N., Shakuf, V., Rudzicz, F., & van Lieshout, P. H. (2016). Prosody and Semantics Are Separate but Not Separable Channels in the Perception of Emotional Speech: Test for Rating of Emotions in Speech. *Journal of Speech, Language, and Hearing Research*, 59(1), 72-89.

Ben-David, B. M., & Icht, M. (2017). Oral-diadochokinetic rates for Hebrew-speaking healthy aging population: Non-word vs. real-word repetition. *International Journal of Language & Communication Disorders*, 52(3), 301-310.

Ben-David, B. M., & Icht, M. (2018). The effect of practice and visual feedback on oral-diadochokinetic rates for younger and older adults. *Language and Speech*, 61(1), 113-134.

Ben-David, B.M., Icht, M., Mentzel1, M., Gilad, M., Dor, Y.I., Ben-David, S., Carl, M., and Shakuf, V. (in press). *Challenges and opportunities for telehealth assessment during COVID-19: iT-RES, adapting a remote version of Test of Rating of Emotions in Speech. International Journal of Audiology*.

Davis, A. (1995). *Hearing in Adults*. London:Whurr.

Dupuis, K., & Pichora-Fuller, M. K. (2010). Use of affective prosody by young and older adults. *Psychology and aging*, 25(1), 16.

Engdahl, B., Tambs, K., Borchgrevink, H. M., & Hoffman, H. J. (2005). Screened and unscreened hearing threshold levels for the adult population: Results from the Nord-Trøndelag Hearing Loss Study. *International journal of audiology*, 44(4), 213-230.

Filippi, P., Ocklenburg, S., Bowling, D. L., Heege, L., Güntürkün, O., Newen, A., & de Boer, B. (2017). More than words (and faces): evidence for a Stroop effect of prosody in emotion word processing. *Cognition and Emotion*, 31(5), 879-891.

Fontan, L., Ferrané, I., Farinas, J., Pinquier, J., Tardieu, J., Magnen, C., ... & Füllgrabe, C. (2017). Automatic speech recognition predicts speech intelligibility and comprehension for listeners with simulated age-related hearing loss. *Journal of Speech, Language, and Hearing Research*, 60(9), 2394-2405.

Füllgrabe, C. (2020). On the possible overestimation of cognitive decline: The impact of age-related hearing loss on cognitive-test performance. *Frontiers in Neuroscience*.

Gordon-Salant, S., Frisina, R. D., Fay, R. R., & Popper, A. (Eds.). (2010). *The aging auditory system* (Vol. 34). Springer Science & Business Media.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, 22, 193–225.

Lin, Y., Ding, H., & Zhang, Y. (2020). Prosody dominates over semantics in emotion word processing: Evidence from cross-channel and cross-modal Stroop effects. *Journal of Speech, Language, and Hearing Research*, 63(3), 896-912.

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: a strong connection. *Psychology and aging*, 9(3), 339.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour research and therapy*, 33(3), 335-343.

Oron, Y., Levy, O., Avivi-Reich, M., Goldfarb, A., Handzel, O., Shakuf, V., & Ben-David, B. M. (2019). Tinnitus affects the relative roles of semantics and prosody in the perception of emotions in spoken language. *International journal of audiology*, 1-13.

Pell, M. D., Jaywant, A., Monetta, L., & Kotz, S. A. (2011). Emotional speech processing: Disentangling the effects of prosody and semantic cues. *Cognition & Emotion*, 25(5), 834-853.

Schaette, R., & McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *Journal of Neuroscience*, 31(38), 13452-13457.

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds*.), The handbook of aging and cognition* (p. 155–219). Lawrence Erlbaum Associates Publishers.

Cross-modal Commutativity of Brightness and Loudness Productions

Wolfgang Ellermeier, Florian Kattner, and Anika Raum
Institute of Psychology, Technical University of Darmstadt

Abstract— In their fundamental paper, Luce, Steingrimsson, and Narens (2010, *Psychological Review*, 117, 1247-1258) proposed that ratio productions constituting a generalization of cross-modality matches may be represented on a single scale of subjective intensity, if they meet 'cross-dimensional commutativity'. The present experiment is the first to test this axiom by making truly cross-modal adjustments of the type: "Make the sound three times as loud as the light appears bright!" Twenty participants repeatedly made cross-modal productions of this sort by adjusting the level of a 500-ms burst of noise to result in the desired sensation ratio (e.g. to be three times as intense) compared to the brightness emanating from a grayscale square having sides of 5.7-cm length, simultaneously displayed on the computer screen, and vice versa. Subjects could vary both stimulus dimensions up to approx. two log units above the lowest reference level used: 1.0 to 85 cd/m2 for luminance and 40 to 90 dB(A) for sound pressure level. Cross-dimensional commutativity was tested by comparing a set of fourteen successive 2x3x productions with a set of fourteen 3x2x productions per participant, with each factor involving a mapping from one sensation into the other. When this property was individually evaluated for each of 20 participants and for two possible directions, i.e., starting out with a noise burst or a luminous patch, only 7 of the 40 tests indicated a statistically significant violation (using matched-pairs Wilcoxon signed-rank tests) of cross-modal commutativity. This, provided a few other measurement axioms hold, suggests that both loudness and brightness sensations may be gauged against a common ratio scale of subjective intensity.

Keywords— cross-modal commutativity, brightness perception, loudness perception

Just Noticeable Differences in Neuroscience and Society: The Case of Emil du Bois-Reymond

Gabriel Finkelstein
University of Colorado Denver

Abstract— In 1850 Emil du Bois-Reymond was invited to the Académie des sciences to demonstrate his discovery of what we now call the action current. The French panel of judges refused to recognize his finding, claiming that his methods of experiment, which involved isolated preparations of animal tissue, couldn't be relied on to explain human physiology. Much like du Bois-Reymond, I contend that small differences can indicate larger findings. In my case, differences in the interpretation of scientific experiments can suggest larger differences in values. Historians critical of my method, like the French critics at the Académie, argue that no conclusions can be drawn unless all society anomalies are be accounted for. Here I side with du Bois-Reymond: case studies are to the historian what demonstrations are to the physiologist.

Keywords— just noticeable differences, du Bois-Reymond, history of psychology

Manipulating Self-Efficacy Beliefs in Basketball with Temporal Pressure

Daniel Fortin-Guichard[1,2], Christian Jacques[1], Annie Goulet[1], Simon Grondin1, and Isabelle Giroux[1]
[1]École de psychologie, Université Laval
[2]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam

***Abstract*—** Self-efficacy beliefs (SEB) refer to one's beliefs in his or her capacity to execute a task. Such beliefs are modified according to experiences. The main purpose of this study is to verify if SEB can be manipulated by modulating the temporal pressure for the execution of a task. The task adopted for the study is a sport activity performed by athletes. Twenty-three basketball players were assigned randomly to one of two groups in which SEB for a basketball task was either increased or decreased. During a task limited in time, athletes predicted the number of 3-point baskets they could fare and tried to achieve this number. To increase or decrease SEB, athletes were told they had two minutes to complete the task, but were in reality allocated 15% more or less of this time, respectively. Results show a significant increase or decrease of SEB after the basketball task in line with the intended manipulation. The study shows that unnoticeable changes in the course of time are sufficient to provoke a change in SEB that could affect performance.

***Keywords*—** Temporal pressure, self-efficacy beliefs, sports, basket-ball

I. Introduction

Studies on time perception typically ignore how temporal pressure impacts confidence-related psychological concepts like self-esteem or self-efficacy beliefs (SEB). A recent study shows that temporal pressure does not impact SEB in the context of a non-mastered task (Christandl et al., 2018). What if expertise was necessary? Indeed, SEB refers to one's beliefs in his or her ability to execute a task or a series of specific tasks in order to reach a desired goal (Bandura, 1977). According to Bandura (1986, 1997), SEB are constructed on the basis of four main factors: (a) mastery experiences (e.g., one's past sport performances), (b) vicarious experiences, (c) verbal persuasion and (d) emotional/physiological state. According to Bandura, mastery experiences constitute the most important factor in determining SEB. More precisely, an individual who has had positive experiences (e.g., success) on a given task will tend to have higher and more stable SEB for this task than one who has had negative experiences (e.g., failure). The aim of this study is to show that SEB can be manipulated by a simple manipulation of time.

SEB in sports. SEB is a type of beliefs widely studied in sport psychology because of its known influence on sports performance (e.g., Kane, Zaccaro, & Blair, 1996; Moritz, Feltz, Fahbrach, & Mack, 2000), but is much less frequent in psychophysics. Two sports psychology studies suggest that when an athlete perceives a sport task either as a success or as a failure, his or her SEB toward this task respectively increase or decrease (Fitzsimmons et al., 1991; Gernigon & Delloye, 2003). In these studies, under-performances or over-performances were experimentally induced by manipulating the feedback provided to the athletes following the task. The objective was to determine the effect of this false information on the athletes' SEB in performing the same task in the future. Fitzsimmons and colleagues (1991) informed three groups of 12 male experienced weight lifters that they lifted (a) a greater weight than they did in reality (false positive feedback), (b) a lower weight than they did in reality (false negative feedback), or (c) the exact weight. They measured SEB by asking the athletes which weight they were 100%, 75%, and 50% certain of being able to lift. The results showed an increase in SEB on the same future task when a false positive feedback had been provided for the weight that they were 100% certain of being able to lift, but the results were mitigated for the questions related to 75% and 50% of certainty. Gernigon and Delloye (2003) measured the SEB of sixty-two 60 m sprinters by using the time the athletes thought they would take to sprint and their degree of certainty of achieving this time. The degree of certainty was measured on a scale ranging from 10% (not sure) to 100% (totally sure). Male sprinters who received a faster false time than their actual time increased their SEB for this same future task whereas female sprinters who received a slower time decreased their SEB. As a whole, the results suggested that athletes who received false positive feedback increased their SEB in comparison to those who had received a false negative feedback. Consequently, it is possible to manipulate athletes' SEB on a sport task using false feedback, meaning that mastery experiences on a sport task effectively contribute in determining SEB for the task.

Playing with time. False feedback is a strategy that can only work when it is possible to hide the actual quantification of a performance from the athletes (e.g., not showing their live time to sprinters; Gernigon & Delloye, 2003). However, this deception is not always possible, as some performances in sports can't be hidden. For example, in basketball, athletes can easily count the number of times the ball enters the basket.

A well-known strategy to modulate performance is changing time pressure (Christandl et al., 2018), such as increasing or decreasing the amount of time allowed to complete a sport task; Janelle, 2002; Williams, Singer, & Frehlich, 2002). Although increasing time pressure did not systematically modulate SEB (Christandl et al., 2018), the right amount of pressure could influence SEB. It is difficult to determine, based on the literature on time perception, the amount of acceleration or deceleration required for both keeping time change unnoticeable, and changing performance level. Most literature regarding the just noticeable difference it takes to discriminate time intervals are conducted on brief intervals (Grondin, 2001). The just noticeable difference depends on the magnitude of the intervals under investigation (the idea of the Weber fraction: Grondin, 2014, 2020), on the psychophysical method adopted and on the fact of counting explicitly or not (Grondin, Ouellet,

& Roussel, 2004). Also, the Weber fraction could be close or below 10% if time is marked with auditory signals and above 15% if markers are visual (Grondin, 2003). By generalizing these findings to the overall value of an activity (i.e., very long intervals), it is unlikely that a participant would be able to notice a 15% change in the visual modality. It is possible to detect a change in the value of a second compressed or extended by 15%, but for such a detection to occur, the participant would need to pay specific attention to time, attention being critical in the estimation of time (Casini, Macar, & Grondin, 1992; Grondin & Macar, 1992; Macar, Grondin, & Casini, 1994). Therefore, because completing a sport task (e.g., throwing balls in a basket) requires sustained attention, the 15% temporal change appears unlikely to be noticed.

The current study.

Since the impact of temporal pressure on SEB has not been studied in the context of expertise, the main objective of this study is to verify the effect on athletes' SEB of a temporal manipulation while performing a basketball task, namely throws from the 3-point line. It is expected that accelerating time will result in a decrease of SEB, and decelerating time will lead to an increase of SEB.

II. Methods

Participants

To be eligible, participants had to: (a) be 18 years of age or older, (b) play basketball at a college or university level, (c) not have any current medical contraindications for playing basketball, and (d) know the 3-point shootout of the NBA all-stars game.

Following approval from the Université Laval research ethics committee, an agreement was established with seven colleges and one university in the Quebec City region in order to recruit their basketball athletes. The first author attended a practice to falsely inform the athletes of a study about basketball performances. Sixty-three basketball athletes were interested and 40 were excluded because they did not meet inclusion criteria.

Twenty-three athletes (39.1% women; M_{age} = 19.04 years, SD = 1.33) were randomly distributed (gender taken into account) between two basketball task conditions: decrease of SEB (DSEB; n = 11) and increase of SEB (ISEB; n = 12). Table 1 presents the sociodemographic data according to conditions, and no significant differences are observed between conditions on the variables.

Material

QMinim Online Minimization (Saghaei & Saghaei, 2011) software was used to randomly and equitably distribute the men and women to conditions. A stopwatch visible from at least a 15 meter distance that ran down 15% too quickly, for the DSEB condition (1min. 42 secs.), and 15% too slowly, for the ISEB condition (2mins. 18 secs.), was used for the basketball task. The required visibility of the stopwatch was determined based on the typical dimensions of a basketball court, where the basket is 7.24 m away from the 3-points line and we wanted to make sure the stopwatch was always visible by roughly doubling this distance. A basketball court, 25 men's basketballs (between 23.8 and 24.8 cm diameter), 25 women's basketballs (between 23 and 23.5 cm in diameter), and an 18-inch diameter basketball hoop set at 3.05 meters high (International Basketball Federation, 2012) were also used.

Instruments

Eligibility questionnaire. This homemade telephone administered questionnaire includes 23 questions (12 short answers, nine dichotomous answers, and two using a scale ranging from -5 [Less than my peers] to 5 [More than my peers]) to determine athletes' eligibility for the study. To avoid cuing the athletes on the basketball task they would perform during the experimentation, the question specifically related to this task was diluted throughout other irrelevant questions regarding their basketball performances (e.g., "What is your free throw efficacy, in percentages?"). Sociodemographic questionnaire. This telephone-administered questionnaire created for this study has four open-ended questions regarding the participant's age, gender, mailing address and main position as a basketball player. SEB for a basketball task questionnaire. This homemade questionnaire is inspired by that of Gernigon and Delloye (2003). Two questions assessing the number of baskets (out of 25) that the athlete thinks he could fare and his degree of certainty on a scale from 10% (not sure) to 100% (totally sure) were verbally administered twice: (1) Before the basketball task and (2) After the basketball task: Same questions, but regarding a hypothetical future basketball task. A sample item would be: "Out of 25 3-point shots, how many you think you can fare?"

Table 1
Sociodemographic characteristics of the athletes according to experimental condition and for the overall sample.

Variables	DSEB (n = 11)	ISEB (n = 12)	Total (N = 23)
Mean age	19.27	18.83	19.04
(SD)	(1.49)	(1.19)	(1.33)
% women	45.5	33.3	39.1
Main position played %			
Wing	54.6	83.3	69.6
Guard	36.4	0.0	17.4
Center	9.1	16.7	13.0

Note. SD = Standard deviation.

Procedure

The first author called interested athletes to administer the eligibility and sociodemographic questionnaires. After the interview, the researcher sent an initial consent form by e-mail to eligible athletes. An appointment was made with the athletes to meet at a gymnasium and conduct the study. During the appointments, the researcher explained to the athletes that they would execute a basketball task similar to the 3-point contest conducted at the NBA all-stars game, which represents a training task commonly carried out by basketball athletes during their practices (validity of this information and fidelity of the task were verified by a recently retired collegiate Division 1 player and a seasoned collegiate Division 2 coach). The researcher explained to the athletes that they would have two minutes to complete the basketball task. In reality, the time on the stopwatch presented to the athletes ran down more or less quickly, depending on the condition. No other clock was visible inside the gymnasium and participant's watch was withdrawn if necessary. The task required that the athletes throw five basketballs from five positions on the 3-point line of a basketball court (see Figure 1) within a duration they were told was two minutes.

A 10 CAD monetary compensation was given for participation. A debriefing was conducted by phone and e-mail when all athletes from a same institution had completed the experiment in order to inform about the real goals of the study and verify the validity of the manipulation. The researcher presented the real goals of the study (i.e., measure of SEB, not performance), the reasons for the dupery (i.e., avoid cuing the athletes on their SEB, making sure they concentrate on the basketball task), and the way that the dupery was conducted (i.e., 15% more or less time to complete the basketball task). Also, athletes were asked if they noticed an acceleration or a deceleration in time, compared to real time, during the basketball task. The researcher sent a post facto consent form by e-mail. All athletes agreed to stay in the study after post facto consent and reported not having noticed the time change.

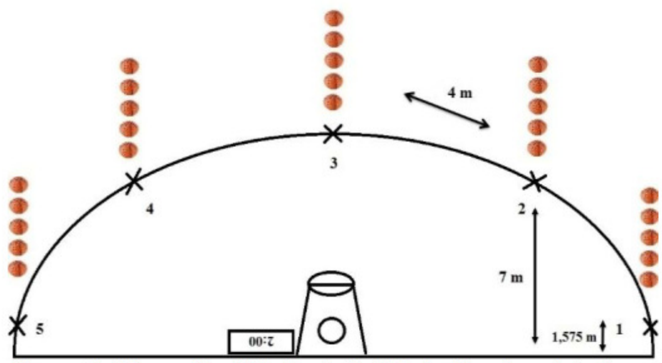

Figure 1. Schematic of the basketball task. Numbers 1 to 5 indicate the order in which athletes had to throw the balls.

Study design

The independent variable corresponds to conditions (DSEB or ISEB). The dependent variable is the SEB for the sports task as measured by the number of basket (out of 25) that the athletes thought they could fare in two minutes. The design of the study is experimental because a manipulation was conducted on the independent variable and the participants were randomly distributed to the conditions (Kirk, 2009).

Statistical analyses

The analyses were conducted using IBM SPSS statistics version 22 software. The sample description was made using means, standard deviations and percentages. The test for the experimental manipulation of the basketball task was conducted using a mixed factorial ANOVA (Tabachnick & Fidell, 2007). This analysis was conducted to test for the presence of a statistically significant change in SEB before and after the basketball task according to condition. Since there are only two measurement times for the repeated measures, the sphericity of the mixed ANOVA is considered to have been met (i.e., no test was run). Normality (i.e., skewness and kurtosis) of the distribution within each group was assessed and considered appropriate.

III. Results

Mean SEB before the basketball task was 9.82 and 10.58 (i.e., numbers of basket they thought they could fare) for athletes in the DSEB and ISEB conditions, respectively. After the task, means were 8.27 and 13.17. Results of the mixed factorial ANOVA revealed that the main effect of measurement time (pre and post task) was not statistically significant, $F(1, 21) = 2.25$, unilateral $p = .07$, $\eta^2_p = .10$, and that the main effect of condition was statistically significant, $F(1, 21) = 3.54$, unilateral $p = .04$, $\eta^2_p = .14$. A significant Time × Condition interaction effect for SEB for the basketball task was found, $F(1, 21) = 35.64$, $p < .01$, $\eta^2_p = .63$. Figure 2 illustrates these effects as well as standard deviations and standard error of means.

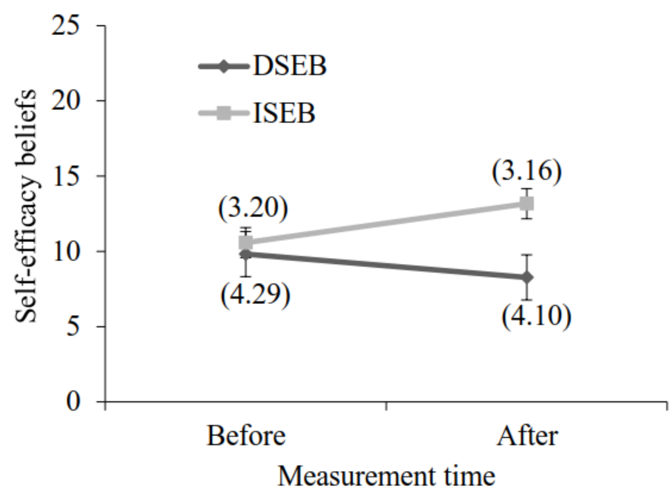

Figure 2. Mean self-efficacy belief (SEB) before and after the basketball task in each experimental condition. The bars represent the standard error of the means and the parentheses represent the standard deviation.

Two repeated-measures ANOVAs were conducted to breakdown the interaction effect and to determine where the differences between measurement times and conditions were. Both ANOVAs revealed the presence of a statistically significant difference between SEB before and after the basketball task: in the DSEB condition, decrease of SEB ($F[1, 10] = 55.58$, $p < .01$, $\eta^2_p = .85$) and in the ISEB condition, increase of SEB ($F[1, 11] = 16.65$, $p < .01$, $\eta^2_p = .60$). Moreover, two independent sample Student t tests were conducted to compare SEB between conditions before and after the manipulation. Before the manipulation, the analyses did not reveal any statistically significant difference between conditions, $t(21) = -0.49$, $p = .63$, $d = 0.21$, while after the manipulation, the difference was statistically significant, $t(21) = -3.22$, $p < .01$, $d = -1.4$. SEB were higher after manipulation in the ISEB condition than in the DSEB condition.

IV. DISCUSSION

The main objective of this study was to verify the impact of increased or decreased course of time on the SEB of basketball players asked to perform a series of throws from the 3-point line.

Results show that it is possible to manipulate SEB of basketball athletes by adding unnoticeable changes (acceleration or deceleration) of the time course during a basketball task. This finding concurs with those observed previously in the literature among athletes participating in sports where actual performance can be hidden from the participant (Fitzsimmons et al., 1991; Gernigon & Delloye, 2003), but contradicts those found in students who solved anagrams (Christandl et al., 2018). Moreover, this finding concurs with the widespread literature in sports science stating that time pressure alters performance (Janelle, 2002; Williams et al., 2002). Athletes probably self-identify to a sport task (either from expertise or enjoyment), at least more so than students performing a non-mastered task (Christandl et al., 2018), potentially explaining why manipulation worked in the present study. The novelty of the present findings reside in the possibility of manipulating SEB (rather than only performance) of athletes participating in sports where they have access to their actual performance with a manipulation of time.

The study shows that athletes involved in their sport activity will not notice a 15% change in time, either acceleration or deceleration, relative to the normal flow of time. The attention of the athletes was on the basketball task, which made it almost impossible to notice that the elapsed time was going too fast or too slow. As noted earlier, perceived duration depends on the attention to time (Casini et al., 1992; Grondin & Macar, 1992; Macar et al., 1994). Indeed, even when their attention is on the task, humans that try to produce or adjust a 1-sec interval could be very imprecise in their estimates (Grondin et al., 2020).

This study has some methodological limitations. First, the relatively small sample size does not provide a high statistical power and raises doubts about the representativeness of the results to all basketball athletes. Second, the use of the degree of certainty based on several performances, as suggested by Bandura (2006), rather than only one actual performance as a measure of SEB, may have been more adequate to measure SEB. Finally, the absence of a control condition undergoing no experimental manipulation leaves the present study without a baseline level for this measure. It should however be noted that the experimental conditions are considered to be mutual comparison conditions.

The current study also has some methodological strengths. The experimental design in which potentially noxious variables were statistically and methodologically controlled for (e.g., gender, clocks hidden in the gymnasium) ensures good internal validity of the study. Also, the precaution taken to ensure the representativeness of the basketball task increases the external validity of the study.

Findings from this study open the way to several questions that could inspire future research. For example, sport psychology researchers should look at the impacts of under-performances and over-performances over the long term on athletes' SEB for several sport-related tasks. Such studies could help determine if athletes' SEB can be modulated only for their sport (higher or lower, like in this study), or if other impacts are observable over the long term (e.g., desire to pursue sports, SEB towards sports in general). From a more methodological point of view, replicating the present study with measures of SEB that follow recommendations made by Bandura (2006) in his guide to build a SEB scale might confirm or disconfirm the present results using more empirically validated measures. Finally, researchers in psychophysics should determine the threshold from which the speed of the flow of seconds no longer corresponds to the actual flow of seconds.

Ultimately, this study shows that it is possible to manipulate SEB for a basketball task with a manipulation of the course of time. It is indeed the first empirical demonstration of this kind. This study contributed to better understand a process underlying performance, namely SEB, and develop a clearer portrait of the relation between time pressure and sports performance.

V. REFERENCES

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1986). *Social foundation of thought and action: A social cognitive theory*. Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The exercise of control. Freeman.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307-337). Information Age.

Casini, L., Macar, F. & Grondin, S. (1992). Time estimation and attentional sharing. In F. Macar, V. Pouthas & W. Friedman (Eds.). *Time, Action, Cognition: Towards Bridging the Gap* (pp. 177-180). Dordrecht, Netherlands: Kluwer.

Christandl, F., Mierke, K., & Peifer, C. (2018). Time flows: Manipulations of subjective time progression affect recalled flow and performance in a subsequent task. *Journal of Experimental Social Psychology*, 74, 246–256. https://doi.org/10.1016/j.jesp.2017.09.015

Fitzsimmons, P., Landers, D., Thomas, J., & van der Mars, H. (1991). Does self-efficacy predict performance in experienced weightlifters? *Research Quarterly for Exercise and Sport*, 62(4), 424-431. https:/doi.org/10.1080/02701367.1991.10607544

Gernigon, C., & Delloye, J.-B. (2003). Self-efficacy, causal attribution, and track athletic performance following unexpected success or failure among elite sprinters. *The Sport Psychologist*, 17(1), 55-76. https:/doi.org/10.1123/tsp.17.1.55

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin*, 127(1), 22-44. https:/doi.org/10.1037/0033-2909.127.1.22

Grondin, S. (2003). Sensory modalities and temporal processing. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 61-77). Hogrefe & Huber Publishers.

Grondin, S. (2014). About the (non)scalar property for time perception. In H. Merchant & V. de Lafuente (Eds) *Neurobiology of interval timing* (pp. 17-32) New York: Springer.

Grondin, S. (2020). *The Perception of Time – Your Questions Answered*. New York: Routledge.

Grondin, S., Laflamme, V., & Tétreault, E. (2020). One psychological second does not necessarily last 1000 ms. *PsyCh journal*, 9(3), 414–416. https://doi.org/10.1002/pchj.323

Grondin, S., & Macar, F. (1992). Dividing attention between temporal and nontemporal tasks: A Performance Operating Characteristic - POC - analysis. In: F. Macar, V. Pouthas and W.J. Friedman (Eds.), *Time, action and cognition: Towards bridging the gap* (pp. 119-128). Kluwer.

Grondin, S., Ouellet, B., & Roussel, M-E. (2004). Benefits and limits of explicit counting for discriminating temporal intervals. *Canadian Journal of Experimental Psychology*, 58(1), 1-12. https:/doi.org/10.1037/h0087436

International Basketball Federation. (2012*). Official basketball rules 2012. FIBA Central Board*. Retrieved October 23rd, 2014, from http://www.fiba.com/downloads/Rules/2012/OfficialBasketballRules2012.pdf

Janelle, C. M. (2002). Anxiety, arousal and visual attention: A mechanistic account of performance variability. *Journal of Sports Sciences*, 20(3), 237–251. https:/doi.org/10.1080/026404102317284790

Kane, T., Marks, M., Zaccaro, S., & Blair, V. (1996). Self-efficacy, personal goals, and wrestlers' self-regulation. *Journal of Sport and Exercise Psychology,* 18(1), 36-48. https://doi.org/10.1123/jsep.18.1.36

Kirk, R. (2009). Experimental design. In R. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology* (pp. 23-45). SAGE.

Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition*, 22(6), 673-686. https://doi.org/10.3758/bf03209252

Moritz, S., Feltz, D., Fahbrach, K., & Mack, D. (2000). The relation of self-efficacy measures to sport performance: A meta-analytic review. *Research Quarterly for Exercise and Sport*, 71(3), 280-294. https://doi.org/10.1080/02701367.2000.10608908

Saghaei, M., & Saghaei, S. (2011). Implementation of an open-source customizable minimization program for allocation of patients to parallel groups in clinical trials. *Journal of Biomedical Science and Engineering,* 4(11), 734-739. https:/doi.org/10.4236/jbise.2011.411090.

Tabachnick, B., & Fidell, L. (2007). *Experimental design using ANOVA*. Duxbury.

Williams, M., Singer, R., & Frehlich, S. (2002). Quiet eye duration, expertise, and task complexity in near and far aiming tasks. *Journal of Motor Behavior*, 34(2), 197-207. https:/doi.org/10.1080/00222890209601941

Comparing Paired Comparison

Stephen W. Link
Department of Psychology, University of California
slink@ucsd.edu

***Abstract*—** The Wave Theory of Difference and Similarity provides equations revealing some unexpected methods of analyzing paired comparison data. The basic equations are examined and then slightly altered to take into account known but unexpected experimental results, such as the "Constant error". Both probabilities and response times are related and illustrate often observed phenomena.

***Keywords*—** paired comparison, Wave Theory, constant error

VI. Introduction

The recent republication of The wave theory of difference and similarity (2020, 1992) in Routledge's "Psychological Revivals" of 20th Century books brought to mind again the amazing work of Gustav Fechner. His creation of the first method of measuring a mental value was not just one of the great scientific accomplishments of the 19th century. It also provided the basis for what is known today as Signal Detection Theory, Ideal Observer Theory, and the methods of testing a null hypothesis. What an achievement! 1860!

Wave theory is an extension of the original ideas developed by Fechner (1860). Those ideas, based on the assumption that Gaussian error clouds measurement, were replaced by Wave theory's stochastic process that accumulates stimulus differences over time intervals until a large value A, or a smaller value –A, of accumulation is first reached. Then a decision occurs. The value of A is under control of the decision maker, larger values lead to better performance but longer times to a decision. The theory predicts response probabilities, response times and relations between these common measures of performance. Because the assumption of Gaussian error is not used, the theoretical predictions about relations among paired comparisons are quite different. This new view allows for a new examination of previous work, even that of Fechner.

Before introducing one of Fechner's classic experiments as a basis for this example consider the Wave theory prediction regarding the judgment of whether a Comparison stimulus, C, is greater or smaller than a Standard, S. The Comparison stimulus is characterized as a Poisson random variable with parameter λ and the Standard as a Poisson random variable with parameter μ. Theoretical derivations prove that when C>S the probability the Comparison is judged to be greater, larger, more than the Standard when the Standard is presented first and the Comparison second for judgment is

$$P(C > S) = \frac{1}{1 + e^{-\theta A}},$$

where $\theta = \ln\left(\frac{\lambda}{\mu}\right)$.

This is known as the Logistic equation. This development from a stochastic theory of comparing differences provided a surprising derivation based on cause and effect rather than its initial creation for curve fitting. The exponent of e is -θA. As A, the amount of accumulated difference increases the probability P(C>S) increases. Requiring more accumulation of comparative difference produces a higher response probability and a longer time to a decision. The value of θ equals ln(λ/μ), the natural logarithm of the ratio of internal signal strengths, λ, for the Comparison, and μ for the Standard μ. As the ratio λ/μ increases the probability of judging C > S increases.

This equation is the basis for a new look at Fechner's experiment described in Elemente (1966/1860 p. 161). Using his own judgments, made in sessions of 64 judgments between the Standard and a Comparison, Fechner produced a table of results for 24,576 judgments. The summary results in terms of correctly judging the Comparison to be greater than the Standard are in Table 1. Standards ranged from 300g to 3000g each with two Comparisons either 1.04 or 1.08 greater than the Standard.

Notice that the probabilities steadily increase as the Standards increase. This invalidates the Weber Law prediction that when stimuli are multiplied by a fixed constant, such as 1.04 or 1.08, the probability of judging the Comparison to be larger than the Standard will remain constant.

Yet there is a constancy that Fechner could not observe because it is hidden within the Wave theory equations that predict the probability of a correct response. Each prediction is based on values of θA in Equation (1) that are estimated by

$$\theta A = ln\frac{P(C > S)}{1 - P(C > S)}.$$

For example, the probability correct for the first value in Table 1 is 0.6597. The corresponding value of θA is 0.662 for comparing the 312g weight against the 300g Standard. Values for each response probability appear in the columns of Table 1 for each Standard and Comparison. As is easily seen for both the 1.04 and 1.08 Comparisons the probabilities and θA values steadily increase as the Standards increase. So where is the constancy?

The values of θ are based on the internal representation of the external stimulus's effect. For the Poisson variables these values are, μ for the Standard and λ1 and λ2 for the 1.04 and 1.08 Comparisons respectively. Comparing θA values for a fixed standard and the two comparisons gives:

$$\frac{A\theta_1}{A\theta_2} = \frac{A}{A}\frac{ln\frac{\lambda_1}{\mu}}{ln\frac{\lambda_2}{\mu}}.$$

TABLE 1

Results from Fechner's one-handed experiment n = 24576
Probabilities: Comparison judged > Standard

STANDARD		C_1=1.04*S			C_2=1.08*S		
	C_1	$P(C_1>S)$	$\theta_1 A$	C_2	$P(C_2>S)$	$\theta_2 A$	$\theta_1 A/\theta_2 A$
300	312	0.660	0.662	324	0.726	0.975	0.679
500	520	0.661	0.668	540	0.764	1.173	0.570
1000	1040	0.702	0.858	1080	0.784	1.290	0.665
1500	1560	0.715	0.921	1560	0.824	1.542	0.598
2000	2080	0.724	0.963	2080	0.813	1.473	0.654
3000	3120	0.713	0.909	3120	0.843	1.679	0.542

The internal values for μ, λ1, and λ2 are unknown. Yet, examining this ratio shows that the ratio equals

$$\frac{A\theta_1}{A\theta_2} = \frac{\ln \lambda_1}{\ln \lambda_2}$$
= CONSTANT

if Weber's Law holds across the increases in values of Standards and Comparisons.

In Figure 1 the ratios of θA values are shown as a function of the Standard weights. Quite clearly the ratios are stable across the Standards, although the largest weight has the smallest value. This is unexpected constancy because it suggests that Weber's Law actually applies to these data. If so what accounts for the steady changes in response probabilities as the Stimulus magnitudes increase?

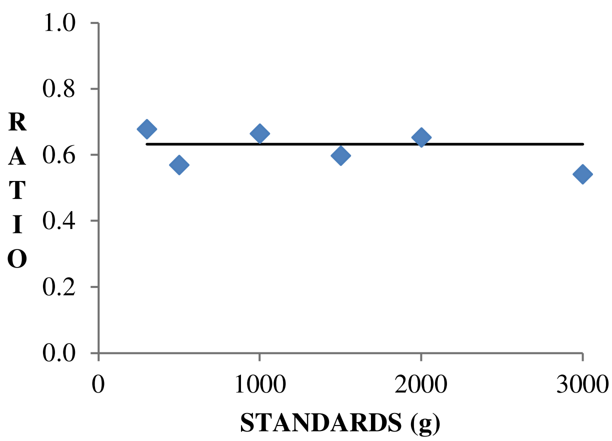

Figure 1. θA Ratios across standards.

If the Comparison stimuli were similarity transforms of the physical stimulus, a reasonable assumption for these weights, then λ1 = k(1.04*S) and λ2 = k(1.08*S). The ratio equals ln(1.04)/ln(1.08)=0.510. the average of the ratios in the last column of Table 1 is 0.618. This value is 20% greater than what would be expected on the basis of a similarity transform of the Standards. Yet the seeming constancy of this ratio, shown in Figure 1 gives rise to the question, Given these ratios are constant, and Weber's Law applies, what causes the increases in response probabilities.

This is where the role of the accumulation of comparative differences shows its effects. Although the constancy of the ratio of θ values suggests the comparison stimuli remain in the same relation across the increases in Standards, the values of A cancel out in the ratio. Assuming these values to be equal for both comparisons may be a mistake.

Suppose changes in response probabilities were a consequence of the experimental subject, Fechner himself, adjusting the amount of comparative difference needed for a response based on the Standard and Comparison used within a 64 trial session. Then values of A could fluctuate from session to session. The stimuli change from session to session but remain fixed within a session.

Keeping within the bounds of the theory and data, consider that the value of A reflects the difficulty of the discrimination.

The smaller the difference between the Standard and Comparison, the more difficult the judgment. The size of A for the smaller Comparison must be greater than for the larger comparison. This increase in θA values across the Comparisons for a fixed Standard is evident in Table 1. As stimulus differences increase in value the average difference between the Comparison and Standard (λ-μ) also increases. To keep performance from declining the values of A must be increased beyond what they were for smaller values of the Standards.

Thus, there are two factors that affect the values of A that apply within a single session of 64 trials. For smaller Standards the values of A may be smaller than for larger Standards. Second, as stimulus difference increases the value of A may increase to accommodate larger differences in the average comparative differences. These changes in A are under the control of the decision maker.

An illustration of these effects is contained in Table 2. The values of θ1 = ln(1.04*S/S) = 0.039 and θ2=ln(1.08*S/S) = 0.077 remain fixed across the entire experiment of 24,576 comparisons. Given these fixed values the predicted probabilities of the Comparison being judged larger than a fixed Standard are

$$P(C_1 > S) = \frac{1}{(1+e^{-0.039A_1})} \quad \text{and} \quad P(C_2 > S) = \frac{1}{1+e^{-0.077A_2}},$$

for the smaller and larger Comparisons respectively. Notice that the only change in response probability as S changes must be through changes in A1 and A2.

TABLE 2

		COMPARISONS					
		1.04			1.08		
		PROBABILITIES			PROBABILITIES		
STANDARD	A_1	OBTAINED	PREDICTED	A_2	OBTAINED	PREDICTED	A_1/A_2
300	17	0.660	0.661	13	0.726	0.731	1.308
500	18	0.661	0.670	15	0.764	0.760	1.200
1000	22	0.702	0.703	17	0.784	0.787	1.294
1500	24	0.715	0.719	20	0.824	0.823	1.200
2000	25	0.724	0.727	19	0.813	0.812	1.316
3000	23	0.713	0.711	22	0.843	0.845	1.045
AVERAGE	21.5	0.696	0.699	17.7	0.792	0.793	1.217

The ratio θ1/θ2 = 0.039/ 0.077= 0.510. Thus, if the values of A1q and A2 are equal the average of such ratios in from Table 1 should also equal 0.510. This is not the case the average basedon data is 0.618. The ratio θ1/θ2 = .510. Multiplying this value by 1.212 yields the observed average value of 0.618. That is, on the average values of A for the smaller Comparison must be 1.212 times larger than the values for the larger comparison whatever increase in A is required as Standards increase in value.

Table 2 provides estimates of A1 and A2 providing close fits to the observed response probabilities, each case considered separately, without requiring that the ratio for each fixed Standard be 1.212. Only integer values for A are used to provide a fit to the observed data

The fit of the theory to the data is, of course, excellent as shown in Figure 2.

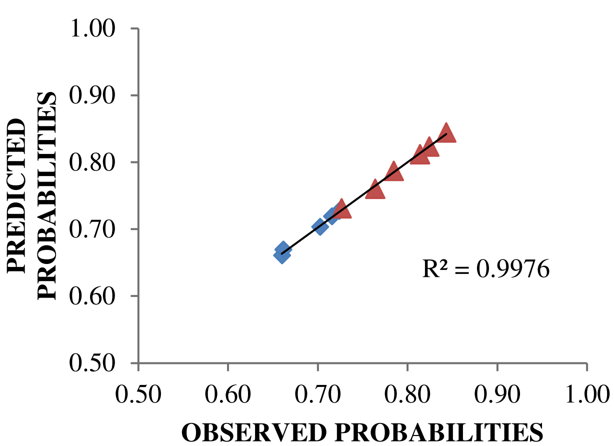

Figure 2. Predicted vs observed response probabilities for 1.04 (diamond) and 1.08 (triangle).

Now, the most important question, given the individual selection of values of A to provide this good fit to the observed probabilities, is the ratio of the average A values at the bottom of Table 2 equal to the actual value of 1.212? The value is 1.217.

In summary, an experiment of Fechner's reported in Elemente (1966/1860) is shown to obey Weber's Law although the data would suggest that Weber's Law did not apply. The Wave theory of difference and similarity analysis shows that Weber's Law does apply. Unknown to Fechner, an amount accumulated stimulus difference is needed for a response to occur. The amount changes according to the difficulties of the judgment. For a fixed Standard, smaller stimulus differences require larger Accumulated differences than for a larger Comparison against the same Standard. This relation remains valid as the magnitudes of Standards increase. As Standards increase larger values of accumulated difference are needed to aid the judgment.

This analysis provides a different interpretation of the reason for the apparent failure of Weber's law suggested by Fechner. I hope this analysis gives Fechner a tickle.

References

Fechner, G. T. (1860) *Elemente der Psychophysics*. Leipzig: Breitkopf & Härtel

Fechner, G. T. (1966) *Elements of Psychophysics*. Vol 1. (E. G. Boring & H. D. H0wes, Eds.: H. E. Adler, Trans). New York: Holt, Rinehart & Winston. (Original work published 1860)

Link, S. W. (1992) *The wave theory of difference and similarity*. Hillsdale, NJ: Lawrence Erlbaum Associates

Link, S. W. (2020) *The wave theory of difference and similarity*. In (Routledge Psychology Revivals) New York: Routledge. (Original work published in 1992)

Temporal Summation for Young and Older Adults

Leah Fostick
Ariel University, Israel

Abstract— Hearing is based on the summation of energy and of changes in energy over time. Temporal summation allows a tradeoff between the duration of sound and its loudness, such that loud sounds can be perceived when they have a short duration, and soft sounds can be perceived when their duration is long enough. Both stimuli and listener characteristics can affect temporal summation: wide-band sounds are perceived with a shorter duration than narrowband, and individuals with a reduction in hearing ability or temporal resolution, such as older adults, require a longer duration to perceive soft or short sounds. The present study was designed to test differences in temporal summation among young and older adults. Older adults show both elevated hearing thresholds and temporal resolution thresholds. Therefore, our prediction was that their detection thresholds will be higher than young adults and that the slope for integrating loudness and duration will be steeper. Forty participants were recruited: twenty young adults (age 20-35 years) with hearing thresholds ≤ 20 dB HL, and 20 older adults (age 65-75) with hearing thresholds ≤ 30 dB HL. The participants performed absolute threshold task for four stimuli (1 kHz, 4 kHz, /e/, /ʃ/) in five durations (5, 20, 50, 100, 200 ms for pure tones and 50, 75, 100, 150, 200 ms for speech sounds). The results showed elevated detection thresholds for older adults, but no difference in slope. There were also effects for spectrum (pure tone vs. speech sounds) and frequency. These results show that as expected, hearing thresholds affect temporal summation, but unexpectedly, that temporal resolution did not affect it.

Keywords— Emotion, Speech perception, Hearing Loss, Older Adults, Spoken Emotions

The Future of Psychophysics: A Personal View

Timothy L. Hubbard[1,2]
[1]Department of Psychology, Arizona State University
[2]College of Doctoral Studies, Grand Canyon University

Abstract— A suggestion that psychophysics is relevant to other research areas begins with a brief autobiographical digression and then discusses possible relevance of psychophysics in studies of imagery, music, neuroscience, embodied cognition, perception and action, naïve physics and perception of causality, and consciousness. Possible issues for a future psychophysics include spatial biases, ecological validity, analogy and metaphor, and central and motor processes. Possible challenges involve recruiting other disciplines, dynamics of stimuli and representations (including memory), complementarity of sensation and judgment, individual and species differences, popularizing psychophysics, and absorption into mainstream psychology.

Keywords— future of psychophysics, interdisciplinary research, challenges for psychophysics, popularizing psychophysics

I. Setting the Stage

When one looks through professional journals, citation indices, academic funding announcements, and college and university catalogs, the word "psychophysics" is not often encountered. This might initially seem surprising, as psychophysics was arguably a very important influence in what would become the scientific study of psychology (e.g., publication of Fechner's Elements of Psychophysics in 1860 predated the founding of the first laboratory for the scientific study of psychology in Leipzig in 1879 by almost two decades). Since the research of Fechner (1860/1966), von Helmholtz (1885/1954), and others in the eighteenth century, psychophysics has made many important contributions to our understanding of human experience and our relationship to the world around us. However, in the past few decades, psychophysics has appeared to diminish in importance within experimental psychology, as discussions of psychophysical topics in many textbooks of *Introductory Psychology and Sensation and Perception* have become increasing smaller. These changes lead to the question: What is the future of psychophysics? Given that I, with some surprise and trepidation, find myself becoming one of the "old guard" in the International Society for Psychophysics, it is perhaps time for me to develop and share some thoughts on the future of our discipline. The tale told here is a highly idiosyncratic one, but perhaps parts of it may resonate with my colleagues and fellow psychophysicists.

II. Imagery and Music (My Early Days)

As a graduate student at Dartmouth College in the late 1980s, I had the good fortune and privilege of working with Prof. John C. Baird, and at that time, one of his interests was in extending psychophysics. I was very interested in imagery, and it seemed a logical step to apply psychophysical techniques developed for the study of perception to the study of imagery. Building on the research of Kosslyn (1980), we conducted several studies examining the portrayal of distance in visual imagery, and we found that the depicted distance to the imaged vanishing point for a given imaged object was a power function of the size of the referent physical object (Hubbard & Baird, 1988) and that addition of "clutter" to a visual image could slightly increase the exponent of this power function (Hubbard & Baird, 1993). These studies eventually led to a more general study of the psychophysics of visual imagery (Baird & Hubbard, 1992), memory psychophysics (Hubbard, 1994), and qualia (Hubbard, 1996). Perhaps as a consequence of this early experience, I have always been mindful that psychophysical methods and concepts can potentially be extended into new areas of research.

Around this time, I also became interested in music perception and cognition, but I only later realized that studies of musical imagery could be viewed as psychophysical studies attempting to link the experienced qualities of musical imagery to the objective qualities of a referent musical stimulus (e.g., such as having participants adjust a metronome to match the tempo of an imaged melody, Halpern, 1988). Although I didn't realize it at the time, my first study in music cognition, in which images of tones or chords were demonstrated to prime perception of subsequently perceived harmonically related tones and chords (Hubbard & Stoeckig, 1988) could be considered an example of psychophysics, as it examined the relationship between the structure of an external stimulus (music) and the structure of the mental representation of that stimulus.

A consideration of music perception and cognition provides interesting boundary conditions for psychophysical approaches. For example, the distinction between prothetic ("how intense") and metathetic ("what kind") stimulus dimensions, while not given much attention in contemporary psychophysics (as most studies tend to focus on prothetic dimensions) is very relevant in music (e.g., loudness is a prothetic dimension, whereas timbre and tone chroma are metathetic dimensions). Also, consideration of musical scales (based on different ratios or temperaments) and psychophysical scales (based on JND units) involve similar issues related to unit size (and perhaps ironically, scales based on musical intervals are not actually interval scales in a psychophysical sense, Hubbard, 2020b). If the results of musical and psychophysical approaches converged (e.g., von Helmholtz, 1885/1954), elegant solutions could emerge. If the results of musical and psychophysical approaches diverged (e.g., the mel scale of pitch, Stevens & Volkmann, 1940, and the helical model of pitch, Shepard, 1982), such divergences could be informative for both domains

III. Relevance to Other Contemporary Research

Just as visual imagery or music might not initially seem to involve psychophysics, I would suggest that there are many Timothy L. Hubbard1,2 1 Department of Psychology, Arizona State University 2 College of Doctoral Studies, Grand Canyon University timothyleehubbard@gmail.com The Future of Psychophysics: A Personal View other areas of contemporary research that might not initially seem to involve psychophysics, but that could be approached with psychophysical methods and concepts. Perhaps the most obvious example is neuroscience. In studies using human participants, many investigators study the effects of lesions or other damage to the nervous system on subjective experience or obtain EEG, PET, fMRI, or other imaging of intact brains that are presented with some stimulus or carry out some task. Such research examines the connections between brain activity and subjective experience, and would seem to fit Fechner's (1860/1966) description of inner psychophysics (i.e., study of the relationship between sensory or other physiological processing and subjective mental experience). Until the development of EEG, PET, fMRI, and other imaging technologies, investigators were not able to study a functioning brain in vivo, and so psychophysics was often limited to what Fechner described as outer psychophysics (i.e., study of the relationship between properties of external stimuli and subjective mental experience). However, the development of brain imaging technologies over the past few decades has enabled the possibility of inner psychophysics. Read (2015) argued that psychophysics has made vital contributions to neuroscience and allows us to interpret physiological measures of neuronal function. Importantly, the increased technological development that allows detailed and precise measures of physiology and physiological processes does not supersede psychophysical methodology; instead, psychophysics complements neuroscience rather than being superseded by neuroscience. Consistent with this, in reviews of brain-imaging studies on musical imagery (Zatorre & Halpern, 2005) and auditory imagery more generally (Hubbard, 2010), many studies were criticized because participants were simply instructed to form an image or the experimental task was presumed to involve imagery. In such cases, given the absence of behavioral or psychophysical evidence that imagery was generated or experienced by the participant, it is not clear whether imagery or some other form of representation was utilized (cf. representational ambiguity, Hubbard, 2018c). Psychophysical (and other behavioral) methods are a critical complement to physiological or neuroscientific investigations. A second example, albeit one related to brain imaging, involves embodied cognition (e.g., Barsalou, 2008; Cox, 2016; Ellis, 2018; Gibbs, 2005). By emphasizing the relationships of subjective experience to sensory responses and to objective properties of physical stimuli, psychophysics is quite salient for the embodied cognition approach in contemporary cognitive psychology. Use of the computer metaphor in early cognitive psychology suggested "mind is to brain as software is to hardware," and this spawned functional approaches in which mental processes were initially thought to be independent of the medium or substrate in which those processes were instantiated (e.g., Block, 1978). Although such approaches were consistent with issues of scaling and threshold addressed in classical psychophysics (i.e., consistent with outer psychophysics), those approaches seemed inconsistent with the lack of concern about the relationship between experiences and the medium or substrate in which those experiences unfolded (i.e., inconsistent with inner psychophysics). In essence, "embodied cognition" seems to be just a new name for inner psychophysics. A third example, albeit one related to embodied cognition, is the focus on links between perception and action (e.g., Prinz, 1997) and on predictive processing (e.g., Hubbard, 2019). This is especially evident in research on spatial biases, as many spatial biases relevant to perception and action are also related psychophysics. An example of this are studies that I and others have conducted on representational momentum (reviewed in Hubbard, 2005, 2018b), momentum-like effects more generally (reviewed in Hubbard, 2015, 2017b), and representational gravity (reviewed in Hubbard, 2020a); even though not initially framed as psychophysics, such biases nonetheless reflect a relationship between properties of invariant physical principles of the environment and objects in that environment (i.e., physics) with properties of our mental representations of the environment and of those objects. More generally, implicit encoding of subjective effects of physical principles on our mental representations of the environment and those objects is a clear example of outer psychophysics. Such studies can also be integrated with existing psychophysical literature (e.g., studies on anisotropies of space attributable to representational momentum, Hubbard & Ruppel, 2018, are related to issues regarding the geometry of visual space, e.g., Wagner, 2006).

A fourth example, albeit one related to perception and action, is naïve physics and the perception of causality. In studies of naïve physics, observers are shown static or dynamic examples of physical systems (e.g., an object dropped from a moving airplane or a pendulum cut at different points along its arc of motion; McCloskey, 1983; Kaiser, Proffitt, Whelan, & Hecht, 1992) and indicate the outcome or whether the depicted motion is correct. In studies of perception of causality, observers are shown dynamic examples of physical systems (e.g., a moving object collides with an initially stationary object that then moves away; Hubbard, 2013; Michotte, 1946/1963) and judge whether the first object caused the motion of the second object. Although psychophysics has traditionally been concerned with the relationship between the objective physical qualities of objects in the world and our sensory response or subjective experience of those stimuli, I would suggest that psychophysics includes our implicit understanding of the physics of the world and that such implicit understanding might be incorporated into our representational system (Hubbard, 2006, 2019).

A fifth example involves consciousness. Although there are some obvious examples of the relevance of psychophysics to consciousness (e.g., psychophysical studies involving absolute thresholds and difference thresholds would clearly fall within the scope of consciousness studies), there are other potential applications (e.g., applications in the study of imagery and memory have already been noted). An interesting possibility is that methods in the study of consciousness that are not considered typical psychophysical methods might be adapted for investigation of psychophysical questions. For example, priming is a commonly used method in research involving conscious or nonconscious processing. In studies of priming, presentation of an initial stimulus (e.g., letter strings, Neely, 1991; visual features of an object, Kristjánsson, Ingvarsdóttir, & Teitsdóttir, 2008; musical chords, Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003) can facilitate processing of a subsequent stimulus and can also influence subsequent social behavior (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012; Smeesters, Wheeler, & Kay, 2010). A related notion is that implicit (i.e., nonconscious) associations can interfere with cognitive processing of stimuli (e.g., Banaji & Greenwald, 2016). Consideration of nonconscious priming and of implicit associations (and of implicit processing in general) might be considered to involve a psychophysics of the unconscious.

IV. A PSYCHOPHYSICS OF THE FUTURE

Westheimer (2015) responded to the "reported death of psychophysics", and he suggested that "psychophysics is not only alive and continuing to serve as an important part of the armamentarium, but its future as an inescapable and permanent ingredient of the enterprise remains unchallenged" (p. 2). To paraphrase Mark Twain, it appears that reports of the death of psychophysics have been greatly exaggerated. Given this, I suspect that a psychophysics of the future would include the related areas of contemporary research discussed in Part III, as well as new areas of research.

There are many spatial biases (for a comprehensive review, see Hubbard, 2018d), some of which have received attention in psychophysical literature (e.g., Gestalt principles of perceptual grouping, various visual illusions). Indeed, there is a history of investigation of spatial biases within psychophysics, with one of Fechner's (1876) first investigations involving the purported preference for a rectangle whose proportions were in the ratio of the golden mean (for consideration of other spatial biases in art and aesthetics, see Hubbard, 2018a). Even so, many spatial biases that were unknown just a few decades ago are now receiving considerable interest from cognitive and perceptual psychologists (e.g., action-specific perception, Witt, 2018). It is possible that study of these more recently-discovered spatial biases could benefit from psychophysical approaches and also contribute to psychophysics. Given that many spatial biases involve dynamic processes, by examining such biases, a psychophysics of the future would focus more on changes and dynamics and less on static representations that are operated upon. Indeed, an increasing emphasis on dynamic possessing is a change that cognitive psychology as a whole has been undergoing (e.g., Freyd, 1987; Schöner, 2008; Ward, 2002).

From its early days until relatively recently, research in psychophysics often involved construction of instruments and other apparati (e.g., see Nicolas & Vobořil, 2019). However, contemporary research often involves stimuli that are computer-generated. One can easily extrapolate from use of computer-generated stimuli to construction of virtual computergenerated environments, and a psychophysics of the future could play an important role in development of virtual reality technologies. A psychophysics of the future will probably continue to rely heavily on computer-generated stimuli. Even so, it will be important to remain mindful of ecological validity and continue to study perception and interaction with physical stimuli in typical everyday 3-D environments. Although not addressing psychophysics per se, Ellis (2018, p. 38) recently noted that "there is much to discover by releasing humans from their chin rests and fixation crosses", and a similar point could be made about psychophysics. Such an idea is also consistent with the embodied cognition notion that observers are not passive recipients of stimulation but are active and mobile organisms who explore and sample their environment (cf. Gibson, 1979; Witt & Riley, 2014).

A psychophysics of the future should address analogy and metaphor. Lakoff and Johnson (1980, 1999) presented persuasive arguments that much of human cognition involves use of analogy and metaphor (e.g., space is a container). Many physical qualities can supply analogies and metaphors for our understanding of physical or imaged stimuli. For example, a melody is objectively just a sequence of notes, but we often metaphorically speak of music as involving motion through space (e.g., melodies move by steps or leaps, melodic contours ascend or descend, the presence of passing or leading tones, etc.), and it has been suggested that music can be understood as involving analogues of physical principles (Hubbard, 2017a; Larson, 2012). A psychophysics of the future could investigate effects of analogical and metaphorical uses of physical principles in additional domains, as well as effects of concrete aspects of physical principles, on mental representation.

Although psychophysics offers much more than thresholds and scaling, many textbook discussions of psychophysics focus on psychophysical methods in the determination of thresholds and scaling. Although there are well-known examples of higherlevel processes involving thresholds (e.g., in signal detection theory), psychophysics is often portrayed as focused primarily on questions regarding sensory input and subjective signal strength. A psychophysics of the future will involve greater extension of psychophysical methods from sensory analysis to central processes (cf. Kaernbach, Schröger, & Müller, 2004), and consistent with this, Barack and Gold (2016, p. 121) note "advances in psychophysical approaches often involve new ways to identify and account effectively for the strategic choices made by subjects." As noted earlier, there have been preliminary extensions from perception to memory in the memory psychophysics literature and in the imagery literature, but this research is still in its early phases. Also, there has been little psychophysical investigation of motor processes (with a notable exception being exertion, e.g., Borg, 1982). Consistent with ideas of perception-and-action and embodied cognition, a psychophysics of the future will extend use of psychophysical methods from sensory analysis to motor processes.

V. GENERAL CHALLENGES

There are a number of opportunities and challenges for a psychophysics of the future. The areas mentioned in Parts III and IV present specific opportunities and challenges, and in addition, there are several other more general challenges.

One challenge involves recruiting investigators from other research areas to the banner of psychophysics. As discussed in Parts III and IV, there are many areas of research in which psychophysical methods and concepts are relevant; however, in the majority of these areas, investigators would probably not self-identify as psychophysicists. As some members of the ISP might recall, in 2014 Prof. Eugene Galanter sent an open letter to the ISP in which he proposed that the society change its name from the International Society for Psychophysics (ISP) to the International Society for Psychophysics, Psychometrics, and Psychobiology (ISPPP). Prof. Galanter's suggestion was based in part on the notion that psychophysics, psychometrics, and psychobiology use similar methods. While I agreed with Prof. Galanter's suggestions that the ISP needed to increase our membership by appealing to colleagues who were using similar methods or investigating similar topics, I was not in favor of changing our name. In a response to Prof. Galanter's letter, I pointed out that I thought of psychobiology as a part of psychophysics (cf. Fechner's inner psychophysics); similarly, I thought of psychometrics as a part of psychophysics, especially when one considered previous attempts by Thurstone, Stevens, and others in applying psychophysical scaling methods to nonsensory issues (e.g., criminal offenses and punishments).

A second challenge involves increasing the focus on dynamic aspects of stimuli and representations. Although some elements of psychophysics involve dynamics (e.g., sequence effects in judgments of intensity, changes in beta and criterion in signal detection theory, extrapolation of location in representational momentum), the role of changes in time is not always clear. There are two domains of particular importance. First, there might be considerable flexibility in the extent of the temporal window of processing (Barack & Gold, 2016). The size of the temporal window over which integration of perceptual information can occur has implications for many perceptual phenomena (e.g., flash-lag effect, Hubbard, 2014), but has not often been widely considered within psychophysics. Second, there has been little investigation of how psychophysical functions might change over time. There are at least two forms of this, and these involve differences between exponents for perception and memory of the same stimulus quality (e.g., Algom, 1992; Petrusic, Baranski, & Kennedy, 1998) and changes in exponent for the memory of a given stimulus quality as a function of latency since perception (Hubbard, 1994). The importance of dynamics is also illustrated in naïve physics, as participants are more likely to provide a correct answer when stimuli are moving (Kaiser et al., 1992), and in perception of causality, in which at least some stimuli are usually in motion.

A third challenge is to clarify differences between sensation and judgment. Baird (1997) pointed out that psychophysical theory often exists in two forms, the first of which focuses on sensory processes and attempts to control or eliminate effects of cognition (e.g., sequence or other context effects), and the second of which focuses on cognitive variables (e.g., judgment strategies) and in which sensory variables have only a minor role. This suggests a type of complementarity in which the same phenomenon can be conceptualized in alternative and apparently contradictory ways. Such a complementarity is similar to the wave-particle duality in quantum physics, but whether such an apparent duality can be resolved in psychophysics or is an unresolvable fundamental characteristic of psychophysics remains to be determined.

A fourth challenge is expanding study of the differences in psychophysical findings. Such an expansion involves studies of differences across humans and studies of differences across species. Psychophysics has traditionally focused on basic fundamental aspects of perception, and a general underlying assumption has been that the studied processes are the same in all humans. Although there have been a small number of studies of individual differences (e.g., Marks, Borg, & Ljunggren, 1983), mainstream psychophysics has typically not considered individual differences. Analogous to the topic of individual differences in humans is the topic of differences across species and how nonhuman animal species differ from each other and from humans. Although there have been many investigations of "animal psychophysics" (e.g., Malone, 2017; Sarris, 2006), there is still much to be learned.

A fifth challenge reflects issues related more to public relations than to science, and those involve how to best increase the exposure of psychophysics in the scientific community and to the broader public. How to publicize what we do as individual scientists and as a discipline? This topic has been discussed, both formally and informally, at previous Fechner Day meetings, but it isn't entirely clear (at least to me) how we should proceed. Perhaps we need an advertising campaign along the lines of "Psychophysics: it's just not for nerds…" or perhaps T-shirts that read "what part of S = kIg or S = (I/k)ln(I/I0) don't you understand?". Although tongue-in-cheek, such an attempt to bridge between psychophysical concepts and popular culture could be useful. Perhaps just as the Association for Psychological Science promises to "give away psychology in the public interest", perhaps we should consider how to best "give away psychophysics in the public interest".

A final challenge is how to keep psychophysics as a distinct and separate subdiscipline and from being completely absorbed into the mainstream of experimental psychology. Although such absorption might be seen as the ultimate success story (in that psychophysical concepts become part of the general paradigm in experimental psychology), it is not entirely clear if such a result is actually desirable. As noted earlier, the founders of psychophysics were among the first to treat psychology as an experimental and quantifiable science, and many elements of psychophysics have already been drawn into the mainstream. Whether or not psychophysics can maintain its own unique identity remains to be seen. In a very real sense, psychophysics could become a victim of its own success.

VI. References

Algom, D. (1992). Memory psychophysics: An examination of its perceptual and cognitive prospects. In D. Algom (Ed.). *Psychophysical approaches to cognition* (pp. 441-513). New York: Elsevier.

Baird, J. C. (1997). *Sensation and judgment: Complementarity theory of psychophysics*. Hillsdale, NJ: Erlbaum.

Baird, J. C., & Hubbard, T. L. (1992). The psychophysics of visual imagery. In D. Algom (Ed.), *Psychophysical approaches to cognition* (pp. 389-440). New York: Elsevier.

Banaji, M. R., & Greenwald, A. G. (2016). *Blindspot: Hidden biases of good people*. New York: Bantam Books.

Barack, D. L., & Gold, J. I. (2016). Temporal trade-offs in psychophysics. *Current Opinion in Neurobiology*, 37, 121-125.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617-645.

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 159-171.

Block, N. (1978). Troubles with functionalism. In N. Block (Ed.) *Readings in philosophy of psychology*, Vol 1. (pp. 268-305). Cambridge, MA: Harvard University Press.

Borg, G. A. V. (1982). Psychophysical bases of perceived exertion. *Medicine & Science in Sports & Exercise*, 14, 377-381.

Cox, A. (2016). *Music and embodied cognition*. Bloomington, IN: Indiana University Press.
Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLOS ONE*, 7(1): e29081.
Ellis, R. (2018). *Bodies and other objects*. New York: Cambridge University Press.
Fechner, G. T. (1966). *Elements of Psychophysics*. New York: Holt, Rinehart and Winston. Original published 1860.
Fechner, G T. (1876). *Vorschule de Aesthetik*. Leipzig: Brietkopt und Hatrtel.
Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427-438.
Gibbs, R. W., Jr. (2005). *Embodiment and cognitive science*. New York: Cambridge University Press.
Gibson, J. J. (1979). *The ecological approach to visual perceptio*n. Boston: Houghton Mifflin.
Halpern, A. R. (1988). Perceived and imaged tempos of familiar songs. *Music Perception*, 6, 193-202.
von Helmholtz, H. (1954). *On the sensations of tone*. New York: Dover Publications. Original published 1885.
Hubbard, T. L. (1994). Memory psychophysics. *Psychological Research/Psychologische Forschung*, 56, 237-250.
Hubbard, T. L. (1996). The importance of a consideration of qualia to imagery and cognition. *Consciousness and Cognition,* 5, 327-358.
Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, 12, 822-851.
Hubbard, T. L. (2006). Bridging the gap: Possible roles and contributions of representational momentum. *Psicologica*, 27, 1-34.
Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin*, 136, 302-329.
Hubbard, T. L. (2013). Phenomenal causality I: Varieties and variables. *Axiomathes*, 23, 1-42.
Hubbard, T. L. (2014). The flash-lag effect and related mislocalizations: Findings, properties, and theories. *Psychological Bulletin*, 140, 308-338.
Hubbard, T. L. (2015). The varieties of momentum-like experience. *Psychological Bulletin*, 141, 1081-1119.
Hubbard, T. L. (2017a). Momentum in music: Musical succession as physical motion. *Psychomusicology: Music, Mind, and Brain*, 27, 14-30.
Hubbard, T. L. (2017b). Toward a general theory of momentum-like effects. *Behavioural Processes*, 141, 50-66.
Hubbard, T. L. (2018a). Aesthetics and preferences in spatial and scene composition. In T. L. Hubbard (Ed.). *Spatial biases in perception and cognition* (pp. 223-240). New York: Cambridge University Press.
Hubbard, T. L. (2018b). Influences on representational momentum. In T. L. Hubbard (Ed.). *Spatial biases in perception and cognition* (pp. 121-138). New York: Cambridge University Press.
Hubbard, T. L. (2018c). Some methodological and conceptual considerations in studies of auditory imagery. *Auditory Perception & Cognition*, 1, 6-41.
Hubbard, T. L. (Ed.) (2018d). *Spatial biases in perception and cognition*. New York: Cambridge University Press.
Hubbard, T. L. (2019). Momentum-like effects and the dynamics of perception, cognition, and action. *Attention, Perception, & Psychophysics*, 81, 2155-2170.
Hubbard, T. L. (2020a). Representational gravity: Empirical findings and theoretical implications. *Psychonomic Bulletin & Review*, 27, 36-55.
Hubbard, T. L. (2020b). *The Pythagorean comma and the preference for stretched octaves*. Submitted for publication.
Hubbard, T. L., & Baird, J. C. (1988). Overflow, first-sight, and vanishing point distances in visual imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 641-649.
Hubbard, T. L., & Baird, J. C. (1993). The effects of size, clutter, and complexity on the vanishing point distance in visual imagery. *Psychological Research/Psychologische Forschung*, 55, 223-236.
Hubbard, T. L., & Ruppel, S. E. (2018). Representational momentum and anisotropies in nearby visual space. *Attention, Perception, & Psychophysics*, 80, 94-105.
Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 656-667.
Kaernbach, C., Schröger, E., & Müller, H. (Eds.) (2004). *Psychophysics beyond sensation*. Mahwah, NJ: Erlbaum.
Kaiser, M. K., Proffitt, D. R., Whelan, S. M., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 669-690.
Kosslyn, S. M. (1980). *Image and mind. Cambridge*, MA: Harvard University Press.
Kristjánsson, Á., Ingvarsdóttir, Á., & Teitsdóttir, U. D. (2008). Object- and feature-based priming in visual search. *Psychonomic Bulletin & Review*, 15, 378-384.
Lakoff, G, & Johnson, M. (1980). Metaphors we live by. Chicago, IL: University of Chicago Press. Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.
Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Bloomington, IN: Indiana University Press.
Malone, J. (2017). Animal psychophysics: The study of sensation in nonverbal organisms. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.). *APA handbook of comparative psychology: Perception, learning, and cognition* (pp. 3-24). Washington, DC: American Psychological Association.
Marks, L. E., Borg, G., & Ljunggren, G. (1983). Individual differences in perceived exertion assessed by two new methods. *Perception & Psychophysics*, 34, 280-288.
McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. L. Stevens (Eds.). *Mental models* (pp. 299-324). Hillsdale, NJ: Erlbaum.
Michotte, A. (1963) *The perception of causality*. New York: Basic Books. Original published 1946.
Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nicolas, S., & Vobořil, D. (2019). Weber's Compass and Aesthesiometers: History of the technical evolution of devices for tactile discrimination. *L'Année Psychologique*, 119, 97- 170.

Petrusic, W. M., Baranski, J. V., & Kennedy, R. (1998). Similarity comparisons with remembered and perceived magnitudes: Memory psychophysics and fundamental measurement. *Memory & Cognition*, 26, 1041-1055.

Prinz, W. (1997). Perception and action planning. *European Journal of Cognitive Psychology*, 9, 129-154.

Read, J. C. A. (2015). The place of human psychophysics in modern neuroscience. *Neuroscience*, 296, 116-129.

Sarris, V. (2006). *Relational psychophysics in humans and animals.* New York: Psychology Press/Taylor & Francis.

Schöner, G. (2008). Dynamical systems approaches to cognition. In R. Sun (Ed.). *The Cambridge handbook of computational psychology* (pp. 101-126). New York: Cambridge University Press.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89, 305-333.

Smeesters, D., Wheeler, S. C., & Kay, A. C. (2010). The role of perceptions of the self, others, and situations in connecting primed constructs to social behavior. *Advances in Experimental Social Psychology*, 42, 259-317.

Stevens. S. S., & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology*, 53, 329-353.

Wagner, M. (2006). *The geometries of visual space*. Mahwah, NJ: Erlbaum.

Ward, L. M. (2002). *Dynamical cognitive science*. Cambridge, MA: MIT Press.

Westheimer, G. (2015). Reported death of psychophysics. *Journal of Vision*, 15(9):17, 1-2.

Witt, J. K. (2018). Spatial biases from action. In T. L. Hubbard (Ed.). *Spatial biases in perception and cognition* (pp. 307-323). New York: Cambridge University Press.

Witt, J K., & Riley, M. A. (2014). Discovering your inner Gibson: Reconciling action-specific and ecological approaches to perception-action. *Psychonomic Bulletin & Review*, 21, 1353-1370.

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and the auditory cortex. *Neuron*, 47, 9-12.

The Pythagorean Comma and the Stretched Octave: A Surprising Similarity

Timothy L. Hubbard[1,2]
[1]Department of Psychology, Arizona State University
[2]College of Doctoral Studies, Grand Canyon University

Abstract— Pythagorean tuning, which derives all musical intervals of a Western chromatic scale from repeated application of a fifth and then a subsequent reduction (collapse) into a single octave, results in an octave interval that is tuned to a ratio slightly larger than 2:1, and the difference between a Pythagorean octave and a 2:1 octave is referred to as the Pythagorean comma. Empirical studies suggest listeners prefer a stretched octave, that is, an octave that is tuned to a ratio slightly larger than 2:1. The magnitude of the Pythagorean comma and the magnitude of the stretch from a 2:1 ratio in preferred octave tuning are similar, and it is hypothesized these phenomena might be connected. Implications and consequences discussed include how the Pythagorean comma predicts preference for a stretched octave, rejection of uncertainty as a cause of preference for a stretched octave, use of hybrid tuning by musicians, how the Pythagorean comma and preference for a stretched octave might be related to tension and musical aesthetics, the nature of "interval" and "scale" in a musical sense and in a psychophysical sense, a role of motion through auditory pitch space in the Pythagorean comma and preference for a stretched octave, and how elements of Pythagorean tuning could be incorporated into representation of musical interval size.

Keywords— *Pythagorean comma, stretched octaves, musical tuning, musical intervals, musical aesthetics*

I. Introduction

The ratios between auditory frequencies of pitches in different musical intervals are important elements of tuning systems, scales, and the specification of musical interval sizes, and there has been debate regarding the merits of different musical tuning systems (e.g., Donahue, 2005; Isacoff, 2003). This article considers the Pythagorean comma and considers evidence that listeners prefer an octave that is slightly stretched (i.e., an octave tuned to a ratio slightly larger than 2:1). Both the Pythagorean comma and the preference for a stretched octave are related to tuning systems used to specify musical interval size, and given that many discussions of different tuning systems are available (e.g., Barbour, 2004; Burns, 1999; Donahue, 2005; Durfee & Colton, 2015; Isacoff, 2003), only aspects of tuning systems relevant to the Pythagorean comma and the preference for a stretched octave are discussed here. A surprising similarity of the ratio of the Pythagorean octave to a 2:1 octave tuning and the ratio of the preferred stretched octave to a 2:1 octave tuning is noted, and this similarity is consistent with a hypothesis that elements of Pythagorean tuning might be incorporated in representation of musical interval. Implications and consequences of this hypothesis are considered.

II. The Pythagorean Comma

The Pythagorean system for deriving musical intervals involves constructing a scale by repeated application of the fifth and the subsequent reduction (i.e., collapsing) of those intervals into a single octave. This is illustrated by the circle of fifths (see Figure 1). Starting at an initial pitch of C, moving up in frequency by a fifth (i.e., a 3:2 ratio of the auditory frequencies comprising the interval) yields a pitch of G. Moving up a fifth from G yields a pitch of D. This process of repeatedly moving up by a fifth until returning to C yields all the notes of the Western chromatic scale (albeit in different octaves, which can then be reduced [i.e., collapsed] into a single octave). Applying a 3:2 frequency ratio to derive each successive position on the circle of fifths spans 7 octaves and yields a frequency 129.746

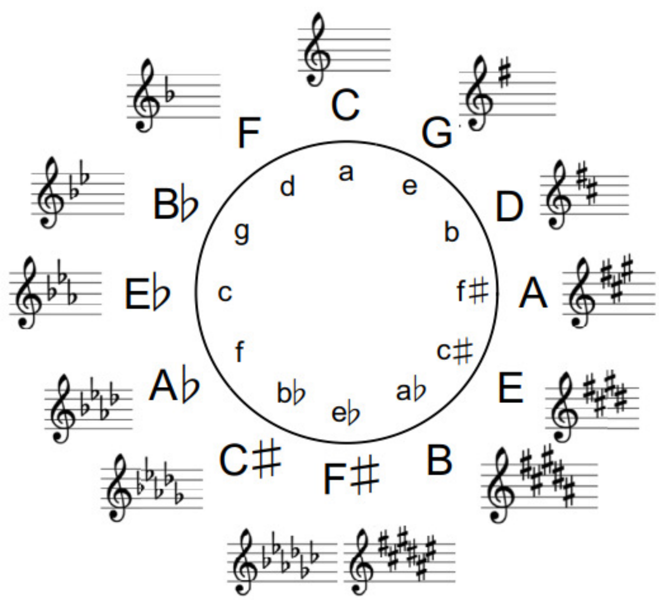

Figure 1. The Circle of Fifths. Upper-case letters indicate major keys, and lower-case letters indicate minor keys. The outermost ring shows the key signature associated with each major key and minor key. Motion clockwise (e.g., C major to G major) involves movement by fifths, and motion counterclockwise (e.g., C major to F major) involves movement by fourths.

([3/2]^12) times that of the starting frequency; however, spanning 7 octaves by repeatedly doubling the frequency of the starting note will yield a frequency 128 (2^7) times that of the starting frequency. The difference between 129.746 and 128 is known as the Pythagorean comma (in music, a "comma" is a pitch interval that results from tuning the same note in different ways). A reduced C tuned by twelve successive applications of a fifth ([3/2]^12) will sound higher in pitch than a reduced C tuned by seven doublings of frequency (2^7). If one instrument was tuned by application of successive fifths and another instrument was tuned by repeated doubling of frequencies, then those instruments would not be in tune with each other across different intervals and tone chroma1 , nor would it be possible to successfully transpose a musical composition into a distant key.

III. PREFERENCE FOR A STRETCHED OCTAVE

One of the most important ratios in music is the 2:1 ratio that defines an octave, and this ratio is present in the music of most, if not all, human cultures (Burns, 1999; Dowling & Harwood, 1986). Given the importance of the octave, it might be predicted that a preference for a 2:1 ratio should be relatively common and strong. However, listeners seem to prefer a slightly "stretched" octave in which the ratio between the two frequencies that form the octave is slightly larger than 2:1. More specifically, when listeners are presented with a musical pitch and adjust a comparison pitch to be an octave higher, they adjust the comparison pitch to be approximately 21 cents sharp (e.g., Ward, 1954; although see Jaatinen, Pätynen, & Alho, 2019) and when judging whether a comparison pitch that varied in frequency around a 2:1 ratio was an accurately tuned octave, they judged a pitch 20 cents sharp to be in-tune (e.g., Dobbins & Cuddy, 1982). Such a pattern has been found in several different musical cultures (for review, see Carterette & Kendall, 1999). Such a judgment does not just affect the octave, but might involve a general stretching of tonal schemata for the entire scale (e.g., Jordan & Shepard, 1987; but see Rosner, 1999). The preference for a stretched octave is sometimes stronger in musicians than in non-musicians (Loosen, 1994, 1995; but see Jaatinen et al., 2019), and this suggests such a preference might reflect musical experience or training. The stretching of the preferred octave size is larger for higher auditorily frequencies (Hartmann, 1993), if the lower pitch of the octave interval is at least 1000 Hz (Sundberg & Lindquist, 1973), and if the pitches are presented in immediate succession rather than separated by silence or by two musically related tones (Cuddy & Dobbins, 1988).

IV. A SURPRISING SIMILARITY

There is a surprising similarity between the Pythagorean comma and the preference for a stretched octave, and this similarity involves the ratio of the Pythagorean octave to a 2:1 ratio and the ratio of the size of the preferred stretched octave to a 2:1 ratio. The ratio of the Pythagorean value of 129.746 ([3/2]^12) and the octave doubling value of 128 (2^7) is 129.746/128 = 1.014. The ratio of the size of the preferred stretched octave (1220 cents) and the size of an octave based on a 2:1 ratio (1200 cents) is 1220/1200 = 1.017. These two ratios of 1.014 and of 1.017 are remarkably similar. Indeed, when the Pythagorean octave is transposed downward by seven (unstretched) octaves, it is 24 cents sharp of the initial (starting) frequency (Donahue, 2005), and this is remarkably similar to the size of the preferred stretched octave2. The Pythagorean comma and the preference for a stretched octave converge on approximately the same size (i.e., the same ratio of auditory frequencies) for the preferred octave tuning.

V. IMPLICATIONS AND CONSEQUENCES

It is possible the Pythagorean comma is unrelated to the preference for a stretched octave and that any similarity in size of the Pythagorean comma and in the stretch of the preferred octave tuning is coincidental. Even so, the similarity is suggestive and consistent with the possibility of a connection3, and so speculation is provided regarding several issues related to the possibility such a connection.

Derivation and Prediction. The preference for a stretched octave is not easily accounted for by appeal to non-Pythagorean tuning systems. However, the preference for a stretched octave can be easily derived from a consideration of Pythagorean tuning and the Pythagorean comma. If interval sizes in the chromatic scale are cognitively specified in terms of successive applications of the fifth and a subsequent reduction, then the resulting Pythagorean comma actually predicts a preference for a slightly stretched octave, that is, the preferred octave size should be a ratio slightly larger than 2:1. Indeed, the similarities of the ratio of a reduced Pythagorean octave to a 2:1 ratio and of the ratio of a stretched octave to a 2:1 ratio are consistent with the hypothesis that encoding of musical intervals is accomplished by mechanisms that are consistent with Pythagorean tuning. Of course, the mere presence of similar ratios and a similar stretching of the preferred octave size is not necessarily indicative of a causal relationship, and additional studies will be required to establish the extent to which representations of interval size incorporate or otherwise involve elements of Pythagorean tuning.

Uncertainty. Judgments of pitch tuning or of musical interval size are relatively imprecise (for reviews, see Burns, 1999; McDermott & Oxenham, 2008), and it might be suggested that such uncertainty in judgments contributes to a preference for a stretched octave. Indeed, melodic intervals as much as 20-25 cents out of tune (based on equal-tempered tuning) are rated as correctly tuned by expert listeners (Vurma & Ross, 2006), and such a deviation from equal-tempered tuning is consistent with Pythagorean tuning and a preference for a stretched octave. Uncertainties in judgments of individual intervals can be summed as intervals are combined (e.g., Ward, 1954, found stretching of a double octave [ratio of 4:1] was approximately twice as large as stretching of a single octave [ratio of 2:1], and this suggests that stretching of successive octaves is additive). However, the argument that a perceived or preferred stretch results from an increase or addition of uncertainties can be rejected, as such uncertainties would be expected to average (cancel) out over repeated additions of intervals. In other words, uncertainty would not favor either a stretching or a compression of interval size, but would result in each interval being added or reduced as equally likely to be represented as larger or as smaller. Any increase in uncertainties regarding interval size would contribute additional variance that would obscure or operate against any systematic bias toward a stretching or compression of interval size (i.e., adding uncertainty would not be expected to produce a bias in a consistent direction such as that demonstrated by the preference for a stretched octave).

Use of Hybrid Tuning. The piano serves as the standard for Western tuning, and as the spectrum for a piano tone is stretched relative to the harmonic series, octaves on the piano are stretched relative to a 2:1 ratio (i.e., the Railsback stretch; Giordano, 2015). Ubiquitous use of the piano as a standard for tuning could certainly contribute to the general preference for a stretched octave. However, tuning a single instrument (e.g., to avoid beating) is not the same as intonation and tuning within an ensemble performance (e.g., Karrick, 1998), as the latter also involves learned sizes of intervals and factors such as expression and voice-leading that might modulate tuning. Studies of intonation have typically not considered effects of a Pythagorean comma (for an exception, see Hellegouarch, 2002, who suggests Pythagorean tuning allows musicians to disambiguate between black-key notes such as D# and Eb, F# and Gb, and C# and Db in a way that not possible within equal-tempered tuning). Consistent with this, string musicians performing music originally composed for equal-tempered tuning often produce music that appears to use a hybrid of just, equal-tempered, and Pythagorean tuning (Borup, 2008; see also Parncutt & Hair, 2018), and it can be speculated that musicians playing other non-fixed pitch instruments, as well as vocalists, might also produce music that would use a hybrid tuning. It might be that a stretched octave is actually the best compromise for such a hybrid tuning.

Tension and Musical Aesthetics. The auditory frequency differences suggested by the ratios of a reduced Pythagorean octave and the preference for a stretched octave to a 2:1 ratio octave could induce a slight tension into the perception of musical intervals. Meyer (1956) suggested that such a tension might be linked to aesthetic qualities of music. If the Pythagorean comma suggests that listeners encode musical intervals based on successive application of the fifth and subsequent reduction into a single octave, then the intervals in music composed or performed using other tuning systems (e.g., equal-tempered) would deviate slightly from the intervals that would be expected (i.e., there would be differences between interval sizes that were expected [based on Pythagorean tuning] and interval sizes that listeners were presented with [based on equal-tempered tuning]), and such deviation might be sufficient to produce the type of tension Meyer linked to aesthetic experience (cf. arousal and aesthetics; Berlyne, 1971). Slight differences between the expected pitch and the perceived pitch due to the Pythagorean comma (and created by harmonics related to Pythagorean tuning) might result in more aesthetic appeal than if such differences weren't present. Relatedly, the fifth (along with the octave) could form part of the structural skeleton underlying a musical composition. The structural skeleton underlying a piece of visual art may play an important role in visual aesthetics of that artwork (e.g., Arnheim, 1974), and an analogous structural skeleton underlying a musical composition might play an important role in the musical aesthetics of that composition.

Musical Scales and Interval Scales. If the preferred size of an octave is stretched slightly beyond a 2:1 ratio, then that raises the question of whether the preferred sizes of other musical intervals might also be stretched by an equivalent amount from the theoretically-specified ratios for those intervals. If other intervals within a musical scale were not stretched by an equivalent amount, then a scale composed of musical intervals would not actually be an interval scale, at least not in the psychophysical sense in which the sizes of intervals between adjacent items are comparable (i.e., the same size). Indeed, other aspects of the scaling of musical stimuli do not correspond to psychophysical scaling (e.g., compare the mel scale of pitch, Stevens & Volkmann, 1940, with the helical model of pitch, Shepard, 1982). However, in considering such issues, two senses of "scale" and "interval" should be distinguished. The first is a musical sense that refers to an ascending or descending sequence of musical notes. In this sense, there can be different interval sizes between different pairs of adjacent notes within the same musical scale (e.g., a major scale involves intervals of 2, 2, 1, 2, 2, 2, and 1 semitone between successive notes). The second sense is a psychophysical sense in which different types of scales involve different relationships between adjacent elements of the scale and the operations that can be performed on those elements (for review, see Baird & Noma, 1978). One type of psychophysical scale is an interval scale, and in this type of scale, the distance between any two adjacent elements (e.g., notes) should equal the distance between any other two adjacent elements (e.g., notes).

As musical scales in the Western tonal tradition are built out of semitones, the question arises as to whether the size of each semitone (e.g., the perceived distance between adjacent pitches in musical pitch space) is the same. Findings of Jordan and Shepard (1987) that tonal schemata are stretched (so that the relative sizes of all intervals were stretched an equal amount) suggest that semitone interval size is preserved across a musical scale, but findings of Rosner (1999) that the preferred tuning of some intervals smaller than an octave involve a ratio smaller than the predicted ratio (i.e., such intervals are compressed) suggest that semitone interval size is not preserved across a musical scale. Perhaps ironically, even though a musical scale is composed of intervals, findings that some intervals are stretched or compressed differently than are other intervals suggest that a musical scale is not an interval scale (Hubbard & Courtney, 2002). Furthermore, scales in major, minor, or other modes in which the distance between adjacent notes can be one, two, or more semitones would not be interval scales in a psychophysical sense. The lack of an inherent interval nature is consistent with findings that individuals differ in the amount of mistuning that is acceptable (e.g., Hall & Hess, 1984) and that preferences in mistuning might result from musical experience (e.g., Loosen, 1994, 1995). Indeed, the ubiquitous use of equal-tempered tuning in Western tonal music amounts to an attempt to impose an interval scale structure (in a chromatic scale, each successive note is 2^[1/12] higher in frequency than the preceding note) on what are actually non-interval scale stimuli.

Movement Through Pitch Space. The ideas of a tension that arises from differences between a perceived and an expected interval size, and that aesthetics might result from such tension or from a desire to return to the tonic or other central pitch, are consistent with the idea of movement through pitch space. Indeed, movement through pitch space underlies the idea of the circle of fifths, and judging interval size would seem to involve movement from one pitch to another through pitch space. If harmonic or melodic intervals are encoded (or otherwise specified) in terms of the distance between the notes of the musical interval in pitch space, then movement through that distance could be implied or implicitly encoded in the representation (cf. Hubbard, 2017; Larson, 2012). Relatedly, previous studies have shown that movement of a stimulus through physical space results in a slight forward (i.e., in the direction of anticipated motion) shift of the judged location of that stimulus (i.e., the judged final location is beyond the actual final location), and this is referred to as *representational* momentum; for review, see Hubbard, 2005, 2018). In judgment of an octave, the actual final location in pitch space of the octave would correspond to a 2:1 frequency ratio to the initial pitch, but the judged final pitch of an octave interval would be shifted slightly forward in the direction of suggested motion in pitch space, thus producing a stretched octave. This stretching is consistent with a slight forward shift in judging a 2:1 ratio distance and with previous findings of a slight forward shift of a moving pitch stimulus in pitch space (e.g., Hubbard & Ruppel, 2013; Johnston & Jones, 2006).

An account of the preference for a stretched octave that emphasizes movement through pitch space (and forward shift of the final pitch) is consistent with an account based on Pythagorean tuning, as the latter explicitly refers to movement around the circle of fifths (and thus through pitch space). However, a motion-based account goes beyond Pythagorean tuning and would presumably influence all musical intervals equally (as representational momentum is generally not influenced by target trajectory length), whereas a Pythagorean tuning account would suggest a greater stretching as more intervals are summed. Additionally, the forward shift from a 2:1 ratio suggested by the Pythagorean comma involves a shift of the actual endpoint of the motion, whereas the forward shift from representational momentum involves a difference between the actual and the judged endpoints. More critically, it is not clear how a motion-based account would explain compression of some intervals (e.g., see Rosner, 1999). Also, in one study memory for the final pitch of a slightly flattened or slightly sharped octave interval was shifted forward or backward, respectively (i.e., toward an in-tune octave) rather than consistently shifted toward a stretched octave (Hubbard, 1993), although stimulus spacing in that study was likely too coarse to detect a preference for a stretched octave. Although preference for a stretched octave is consistent with the Pythagorean comma and with representational momentum from movement through pitch space, the Pythagorean comma and representational momentum do not appear related to each other

Incorporating a Comma in Representation. Representations of pitch and musical interval incorporate information regarding harmonic relationships (e.g., Bharucha, 1987; Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003). A reasonable question to ask is how information regarding a Pythagorean (or other) comma might become encoded or incorporated into a person's representations of pitch and musical interval. Importantly, a comma per se does not have to be explicitly encoded within the representation; rather, it is the subjective effects of a comma on pitch tuning and musical interval perception that would likely be encoded within the representation4. It could be speculated that mappings of different pitch ranges (absolute frequencies) onto the same musical interval representation might tacitly reflect or incorporate consequences of commas (e.g., stretching or compressing); indeed, mapping perceived intervals of the same musical size but composed of different frequencies onto the same musical interval representation is a de facto reduction. All perceived intervals of the same musical size would not map perfectly onto that representation. These failures of stimuli to map perfectly onto the representation of a given interval size could contribute to musical aesthetics (e.g., the differences between equal-tempered intervals and Pythagorean intervals could create tension leading to aesthetic experience) and to the use of hybrid tuning by some musicians. Furthermore, if effects of specific pitch ranges on the general representation of a given musical interval size occurred automatically, then information regarding a comma would not be introspectively available, but would nonetheless influence processing of the same perceived interval in different pitch ranges. However, models of music perception and cognition have not typically addressed the possibility of comma-relevant information in representation.

VI. Summary and Conclusions

The Pythagorean comma is the difference between the tuning of an octave based on repeated application of the fifth and a subsequent reduction and the tuning of an octave based on a 2:1 ratio. The preference for a stretched octave is the tendency for listeners to prefer an octave to be tuned to a ratio slightly larger than 2:1. Interestingly, when the ratio of a reduced Pythagorean octave to a 2:1 ratio octave (1.014) is compared to the ratio of the preferred stretched octave to a 2:1 ratio octave (1.017), the two ratios are very similar, and Pythagorean tuning and the preference for a stretched octave converge to a preferred octave tuning of approximately 20-25 cents larger than an octave based on a 2:1 ratio. Such a similarity is unlikely to result from uncertainties in the encoding of pitch and interval size or in tuning and intonation, as any such uncertainties should average out rather than create a bias (preference) in a specific direction. This similarity is consistent with the possibility of a previously unsuspected connection between the Pythagorean comma and the preference for a stretched octave and with speculation that elements of Pythagorean tuning might be incorporated into the representation of musical intervals. It is perhaps more likely that such incorporation would reflect subjective consequences of commas on subsequent music perception and cognition rather than an objective encoding or descriptive information regarding commas per se, and incorporation of the Pythagorean comma into the representation of musical intervals need not be explicit or available to introspection. Such an incorporation would offer a novel account for the preference for a stretched octave.

There are several directions of research that might result from the hypothesis of a connection between the Pythagorean comma and the preference for a stretched octave. Differences between Pythagorean tuning and the preference for a stretched octave from an (equal-tempered) octave based on a 2:1 ratio might contribute to tension and musical aesthetics. The existence of stretched octaves (in conjunction with previous findings that expert tuning can differ from equal-tempered tuning by a magnitude equal to the size of the Pythagorean comma) is consistent with notion that musicians implicitly use a hybrid tuning in representation of interval size; indeed, a stretched octave might be the best compromise for such a hybrid tuning. The possibility of stretched or compressed intervals suggests that musical scales are not actually interval scales in a psychophysical sense, and distinctions between different meanings of "scale" and "interval" have implications for music theory and for the psychophysics of music. The Pythagorean comma and the preference for a stretched octave might be related to movement through pitch space, and further implications of the dynamics of such movement should be investigated. Although the focus here was on the Pythagorean comma, there are other commas (and perhaps combinations of commas) that might have similar or related effects. Models and theories of representation of pitch and musical interval have not usually considered effects of commas on representation of musical intervals, but the issues discussed here suggest commas might contribute to the representation of pitch and musical interval.

VII. References

Arnheim, R. (1974). Art and visual perception: A psychology of the creative eye (the new version). Berkeley, CA: University of California Press.

Baird, J. C., & Noma, E. (1978). Fundamentals of scaling and psychophysics. New York: Wiley & Sons.

Barbour, J. M. (2004). Tuning and temperament: A short historical survey. New York: Dover Publications.

Berlyne, D. E. (1971). Aesthetics and psychobiology. New York: Appleton-Century-Crofts.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception*, 5(1), 1-30.

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. Journal of Experimental Psychology: *Human Perception and Performance*, 29(1), 159-171.

Borup, H. (2008). Is that in tune, Mr. Mozart? *American String Teacher*, 58, 32-35.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.). The psychology of music, 2nd ed. (pp. 215-264). San Diego, CA: Academic Press.

Carterette, E. C., & Kendall, R. A. (1999). Comparative music perception and cognition. In D. Deutsch (Ed.). *The psychology of music*, 2nd ed.(pp. 725-791). San Diego, CA: Academic Press.

Cuddy, L. L., & Dobbins, P. A. (1988). Octave discrimination: Temporal and contextual effects. *Canadian Acoustics/ Acoustique Cnanadienne*, 16, 3-13.

Dobbins, P. A., & Cuddy, L. L. (1982). Octave discrimination: An experimental confirmation of the "stretched" subjective octave. *Journal of the Acoustical Society of America*, 72, 411-415.

Donahue, T. (2005). *A guide to musical temperament*. Lanham, MD: Scarecrow Press.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. San Diego, CA: Academic Press.

Durfee, D. S., & Colton, J. S. (2015). The physics of musical scales: Theory and experiment. *American Journal of Physics*, 83, 835-842.

Giordano, N. (2015). Explaining the Railsback stretch in terms of the inharmonicity of piano tones and sensory dissonance. *The Journal of the Acoustical Society of America*, 138, 2359-2366.

Hall, D. E., & Hess, J. T. (1984). Perception of musical interval tuning. *Music Perception*, 2, 166-195.

Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *Journal of the Acoustical Society of America*, 93, 3400-3409.

Hellegouarch, Y. (2002). A mathematical interpretation of expressive intonation. In C. P. Bruter (Ed.). Mathematics and Art (pp. 141-148). Berlin, Heidelberg: Springer.

Hubbard, T. L. (1993). Auditory representational momentum: Musical schemata and modularity. *Bulletin of the Psychonomic Society*, 31, 201-204.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, 12, 822-851.

Hubbard, T. L. (2017). Momentum in music: Musical succession as physical motion. *Psychomusicology: Music, Mind, and Brain*, 27, 14-30.

Hubbard, T. L. (2018). Influences on representational momentum. In T. L. Hubbard (Ed.). *Spatial biases in perception and cognition* (pp. 121-138). New York: Cambridge University Press.

Hubbard, T. L., & Courtney, J. R. (2002, November). Does the cognitive representation of melodic interval correspond to an interval scale? *Auditory Perception, Cognition, and Action Meeting*, Kansas City, MO.

Hubbard, T. L., & Ruppel, S. E. (2013). A Fröhlich effect and representational gravity in memory for auditory pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1153-1164.

Isacoff, S. (2003). *Temperament: How music became a battleground for the great minds of Western civilization*. New York: Vintage Books/Random House.

Jaatinen, J., Pätynen., J, & Alho, K. (2019). Octave stretching phenomenon with complex tones of orchestral instruments. *Journal of the Acoustical Society of America*, 146, 3203-3214.

Johnston, H. M., & Jones, M. R. (2006). Higher order pattern structure influences auditory representational momentum. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 2-17.

Jordan, D. S., & Shepard, R. N. (1987). Tonal schemas: Evidence obtained by probing distorted musical scales. *Perception & Psychophysics*, 41, 489-504.

Karrick, B. (1998). An examination of the intonation tendencies of wind instrumentalists based on their performance of selected harmonic musical intervals. *Journal of Research in Music Education*, 46, 112-127.

Larson, S. (2012). *Musical forces: Motion, metaphor, and meaning in music*. Bloomington, IN: Indiana University Press.

Loosen, F. (1994). Tuning of diatonic scales by violinists, pianists, and nonmusicians. *Perception & Psychophysics*, 56, 221-226.

Loosen, F. (1995). The effect of musical experience on the conception of accurate tuning. *Music Perception*, 12, 291-306.

McDermott, J. H., & Oxenham, A. J. (2008). Music perception, pitch, and the auditory system. *Current Opinion in Neurobiology*, 18, 452-463.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago, IL: The University of Chicago Press.

Parncutt, R., & Hair, G. (2018). A psychocultural theory of musical interval: Bye bye Pythagoras. *Music Perception*, 35, 475-501.

Rosner, B. S. (1999). Stretching and compression in the perception of musical intervals. *Music Perception*, 17, 101-113.

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, 89(4), 305-333.

Stevens. S. S., & Volkmann, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology*, 53, 329-353.

Sundberg, J., & Lindquist, J. (1973). Musical octaves and pitch. *Journal of the Acoustical Society of America*, 54, 922-927.

Vurma, A., & Ross, J. (2006). Production and perception of musical intervals. *Music Perception*, 23, 331-344.

Ward, W. D. (1954). Subjective musical pitch. *Journal of the Acoustical Society of America*, 26, 369-380.

ENDNOTES

[1]Although the focus here is on differences between the initial note and the octave, it is not the case that Pythagorean tuning otherwise results in consonant intervals. For example, a major third differs by as much as 20 cents, and the fifth midway between the initial note and octave (e.g., if the sequence around the circle of fifths started on C, then the fifth midway around the circle is F# and corresponds to the tritone of a scale based on the initial note) is often significantly out of tune and referred to as a wolf fifth (e.g., Durfee & Colton, 2015).

[2] It might be argued that reduction should use stretched octaves rather than unstretched octaves. However, as the majority of music composed, performed, and listened to during the past two and a half centuries has been based on equal-tempered tuning involving non-stretched octaves, use of a non-stretched octave is appropriate for specification of intervals, especially as those intervals would be expected to conform to an equal-tempered scale to allow transposition and ensemble performance.

[3] Although the focus here is on relative sizes of the Pythagorean comma and the preferred (stretched) octave tuning, it should be noted that there are several other commas similar in size to the Pythagorean comma, and these include the diaschisma (between 3 octaves and 4 perfect fifths plus 2 major thirds; 19.55 cents), syntonic comma (between 4 perfect fifths and 2 octaves plus 1 major third; 21.51 cents), and septimal comma (between 1 minor seventh and 1 septimal minor seventh; 27.26 cents). Additionally, there are commas smaller than the Pythagorean comma, and these include the septimal kleisma (between 3 major thirds and 1 octave minus 1 septimal comma; 7.71 cents) and kleisma (between 6 minor thirds and 1 octave plus 1 perfect fifth; 8.11 cents), and there are commas larger than the Pythagorean comma, and these include the dieses (between 1 octave and 3 major thirds; 41.06 cents) and greater diesis (between 4 minor thirds and 1 octave; 62.57 cents). Although these other types of commas span different musical interval sizes than does the Pythagorean comma, it is possible that the preference for a stretched octave might be influenced by these other types of commas, or perhaps more likely, that other aspects of pitch and musical interval representation might be influenced by these other commas (or by combined influence of multiple commas) in ways not yet known.

4 An analogous notion is found in the literature on effects of environmentally invariant physical principles on memory for the location of a target. For example, both physical momentum and physical gravity involve mass, but Hubbard (1997, 2020) suggested that representational momentum and representational gravity did not incorporate objective effects of mass per se, but rather reflected the subjective effects of mass for an observer (i.e., effect of mass are subjectively experienced as weight or heaviness along the target axis that is aligned with the direction of gravitational attraction in the external environment). Thus, the representation does not need to reflect the objective comma per se, but can rather reflect the subjective effects of that comma on pitch tuning and musical interval representation.

Open Science ISP Opportunity: Binary Decision Collaboration to Evaluate Information Accrual Models

Diana Kornbrot
University of Hertfordshire, UK
d.e.konrbrto@heerts.ac.uk

Abstract – The open science movement recommends routine sharing, not just of data, but also of *code*. It provides opportunities for collaborations to provide tools to evaluate and estimate key parameters of competing models. This presentation is a call for such a collaboration, led by ISP, to improve our understanding of time and accuracy in information accrual models of binary decision making across domains. Currently, there are two main model classes: the diffusion decision model and the linear ballistic accumulator. Most existing Mss. compares models, often of only one class, for a small set of studies. Consequently, there is minimal consensus about most successful models. Few tools are available that span several model classes and are usable by experimental psychologists with limited mathematical background. Hence the call for collaboration.

Keywords— binary decision, accuracy, reaction time, bias, speed-accuracy.

I. INTRODUCTION

Open Transparent Science is the jumping off point for a proposed collaboration. It ensures transparency and identifies all collaborators and their responsibilities. It identifies the location and ownership of all relevant design decisions, , materials, data, and analyses, Its' practice enables reanalysis of existing data, going well beyond meta-analysis, and there are several recent examples, including (Kornbrot et al., 2018; Open-Science-Collaboration, 2015). It also enables testing of alternative theories on the same data sets, as pioneered by (Dutilh et al., 2018). Here the target is information models of binary decision making (Heathcote et al., 2019; Lindeloev, n.d.; S. W. Link, 1978). The case is made for an ISP led collaboration for this purpose.

II. INFORMATION ACCRUAL & DECISIONS: CURRENT

A. History

There is a long history of information accrual models, including those where ISP members have been pioneers (Laming, 1968; S. W. Link, 1978; Luce, 1963, 1986; Vickers, 1970). This work typically evaluates models using summary statistics including means, variance and skew as well as hits and false alarm rates. Sequential effects are considered seriously in this early work.

B. Current mainstream approaches

There are two main approaches.

The decision diffusion model (DDM) – (Ratcliff, 2018) which assumes that information accumulation has a single drift rate parameter, which is positive for alternative A and negative for alternative B. The linear ballistic accumulator model (LBA) (Heathcote et al., 2019) assumes that there are separate drift rates for each alternative (A,B). Drift rate(s) are determined by difficulty, and so controlled by the experimenter. Respondents can control speed by adjusting the distance between barriers, and response bias by adjusting start point of information accrual. This leads to the common findings that

- Easy tasks are fast AND accurate
- Speed pressure causes a speed-accuracy trade-off
- Favored responses are fast and skewed

Both DDM and LBA assume normal distributions for barriers and drift rates in spite of Link's demonstration of the superiority of the Poisson (Stephen W. Link, 1992). They vary as to assumptions about which parameters are fixed.

C. Model fitting

Models are typically evaluated by goodness of fit measures, and little attention is given to the actual *values* of the parameters. Furthermore, little attention is given to isosensitivity or isobias functions, or to effects of feedback, or to the accuracy of previous trials. Relevant software is complex and rarely fully documented. It is effectively unusable by non-experts. Consequently, many applied psychologists blithely assume that differences in speed imply differences in underlying cognition/perception. The tradition goes back to Freud who attributed delayed responses to underlying conflicts. More recently there is a substantial literature on 'implicit bias that assumes that taking a long time to associate being an engineer with being a woman informs one as to who might make career decisions about some individual. Scant attention is paid to speed-accuracy trade-offs. This is in part because it is so difficult for applied researchers to *apply* information accrual models.

Recently, there have been advances in shared software. Investigators can explore different models for a single data set (Lindeloev, n.d.) and (van den Bergh et al., 2020) provides comprehensive software to find parameters and fit for a variety of information accrual models, see also (Singmann et al., 2014/2020; Wagenmakers, 2008) It is still quite a .lot of work to get parameters for, say 20 participants in 12 conditions, typical of this kind of study. Thus, there is ample room for collaboration for which Dutlilh has set the scene.

III. Dutilh's Collaboration

Dutilh recruited 17 teams from a large number of response time data analysis experts invited by email, see (https://osf.io/9v5gr/). The teams recruited were all indeed experts in the field.

There were 20 student participants who performed a random dot motion detection task (320 trials/condition). There were 12 conditions (2 speed pressures X 2 difficulties X 3 stimulus probabilities).

The analyst team task was to use their chosen model(s) to 'provide for each experiment the inferences about each of four components of response time performance: ease, response caution, bias, and non-decision time'. They were also asked to 'submit a description of their methods. This description of method was asked to meet a reasonable level of reproducibility, describing at least: 1) outlier removal procedures, 2) the applied mathematical model (if one was used), 3) the method of estimation (if applicable), and 4) the rules applied to draw inferences. See (https://osf.io/ktuy7/).

Agreement between teams was generally strong. However, the 17 teams generated 7 different screening criteria, and no pair of teams got identical results, even for the same model. Performance of the two teams that used raw summary means was as good as any of the teams that used models. The specified task did not *require* any parameter estimation, although several methods would have required such parameters to support inferences.

The results were truly impressive. However, most of the teams could not easily provide model parameters on request. Furthermore, there was only one analysis description that enabled *direct* implementation of code to obtain parameters.

In my view this is pioneering work that should lead to further collaborations.

IV. ISP led collaboration proposal

A. Invitation

The first task is to gather to gather an ISP group who would be willing to plan the initiative. They would then prepare the invitation (Dutilh provides a useful template) and an email distribution list including: Dutilh and his co-authors (includes proponent of current models; van den Bergh and co-authors, Lindeloev, Wagenmakers (software producers); ISP, relevant learned societies, Twitter, LinkedIn, etc.

- Title suggestion. Evaluating and implementing models of binary decision making.

ISP team would need to formulate *goals*. A preliminary suggestion comprises evaluating models of information accrual and producing generally usable software to determine parameters. They would also need to identify up to 6 data sets where difficulty, stimulus value (reward or probability), and time pressure are manipulated.

B. Analysts' contribution and reward

Each analyst team will be completely fee to choose their model and method of analysis. They will then produce the following:

- Specification of model, including all parameters
- Method of analysis
- Full input specification
- Full output specification
- Goodness of fit measures
- Instructions for non-specialist users
- Results
- Means and se for all parameter in each condition
- Goodness of fit measure. Chosen by ISP team, e.g. RMS for reaction time (as Lindloev) or WIC.
- Details for supplement
- Tab separated/spreadsheet file with parameters estimates or each participant in each

All analysts would be co-authors of resulting Ms.(s).

V. Collaboration outcome

The outcome would be a major advance in psychological science

- It would identify 'best' models for different domains (memory perception, decisions, etc.)
- It would provide a valuable tool for any researcher (pure or applied) to identify all psychologically relevant parameters for any binary decision task, including supposedly socially relevant 'implicit' bias.

Reactions of ISP members and potential collaborators would be most welcome, d.e.kornbort@herts.ac.uk.

References

Dutilh, G., Annis, J., Brown, S. D., Cassey, P., Evans, N. J., Grasman, R. P. P. P., Hawkins, G. E., Heathcote, A., Holmes, W. R., Krypotos, A.-M., Kupitz, C. N., Leite, F. P., Lerche, V., Lin, Y.-S., Logan, G. D., Palmeri, T. J., Starns, J. J., Trueblood, J. S., van Maanen, L., … Donkin, C. (2018). The Quality of Response Time Data Inference: A Blinded, Collaborative Assessment of the Validity of Cognitive Models. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-017-1417-2

Heathcote, A., Lin, Y.-S., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods*, *51*(2), 961–985. https://doi.org/10.3758/s13428-018-1067-y

Kornbrot, D. E., Wiseman, R., & Georgiou, G. J. (2018). Quality science from quality measurement: The role of measurement type with respect to replication and effect size magnitude in psychological research. *PLoS ONE*, *13*(2), e0192808. https://doi.org/10.1371/journal.pone.0192808

Laming, D. R. J. (1968). *Information theory of choice reaction times*. Academic Press.

Lindeloev. (n.d.). *Reaction Time Distributions: Shiny demo*.

Link, S. W. (1978). The Relative Judgment Theory Analysis of Response Time Deadline Experiments. In *Cognitive Theory* (pp. 117–138). Erlbaum Associates.

Link, Stephen W. (1992). *Wave theory of difference and similarity*. Erlbaum Associates.

Luce, R. D. (1963). Detection and Recognition. In *Handbook of Mathematical Psychology* (pp. 103–189). Wiley.

Luce, R. D. (1986). *Response times*. Clarendon Press.

Open-Science-Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Ratcliff, R. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive Psychology*, *103*, 1–22. https://doi.org/10.1016/j.cogpsych.2018.02.002

Singmann, H., Kalibera, Tomas, & Gretton, M. (2020). *Rtdists/rtdists* [R]. rtdists. https://github.com/rtdists/rtdists (Original work published 2014)

van den Bergh, D., Tuerlinckx, F., & Verdonck, S. (2020). DstarM: an R package for analyzing two-choice reaction time data with the D∗M method. *Behavior Research Methods*, *52*(2), 521–543. https://doi.org/10.3758/s13428-019-01249-7

Vickers, D. (1970). Evidence for an Accumulator Model of Psychophysical Discrimination. *Ergonomics*, *13*(1), 37–58. https://doi.org/10.1080/00140137008931117

Wagenmakers, E.-J. der M. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review*, *15*(6), 1229–1235. https://doi.org/10.3758/pbr.15.6.1229

The Characteristic Properties of Cognitive Processes with Perceptually Integral Stimuli

Yanjun Liu[1] *and James Townsend*[2]
[1]Vanderbilt University
[2]Indiana University

***Abstract*—** How people process multi-dimensional objects is one of essential interests among the field of psychophysics and cognitive psychology. Previous studies showed that in the presence of different sets of dimensions, cognitive systems employed different approaches (i.e., either following a dimensionally distinctive manner or an unitary gestalt manner). In this study, we probed important underlying properties of the cognitive systems that may produce perceptual integrality, applying the sets of methods theoretically derived from general recognition theory (Townsend & Ashby, 1986) and from systems factorial technology (Townsend & Nozawa, 1995). With the utilization of a set of dimensions, namely height and width of rectangles, which were conventionally shown to induce perceptual integrality (i.e., Macmillan & Ornstein, 1998), we examined various types of cognitive independencies including perceptual independence, perceptual separability and decisional separability, and detected fundamental characteristics of the underlying cognitive structures including mental architecture, logical stopping rule and workload capacity. Our results suggested that in the presence of rectangles, the underlying cognitive processes of height and width tended to interact with each other at both perceptual and decisional level. In addition, our results indicated that the cognitive processes of height and width tended to follow a parallel processing manner, and facilitated the processing efficiency of each other. Altogether, the findings of the current study provided strong evidence to support previous findings of perceptually integral processing of rectangular stimuli, and elucidated essential cognitive properties associated with the underlying mechanism of perceptual integrality. Furthermore, the coherent inferences drawn from both theory-driven methods documented a strong potential in the combination of general recognition theory and systems factorial theory.

***Keywords*—** perceptual integration, general recognition theory

Keynote: Research Problems in the History of Psychophysics

David J. Murray
[1]Department of Psychology / Institute for Data Science, Queen's University
murrayd@queens.ca

In July, I felt pleased and honoured to be invited to present a short keynote address at this online ISP meeting for 2020.

I have been writing about the history of psychophysics since the 1980s. Then recently I became aware that I possessed certain 'gap in knowledge' with respect to subtopics in that discipline. Often, these were gaps in my knowledge that could only be filled by turning to the original sources in German or French. Here I wish to select four such 'gaps in knowledge' and describe how my reading of the original sources helped me to reduce those gaps.

First, concerning the beginning of the 19th century, it is well known that Fechner, in his Elements of Psychophysics (1860/1964), adopted Herbart's use of the word "threshold." It is also worth knowing that Herbart, in his Psychology as Science (1824/1890) ascribed "strengths" to individual mental representations involved in conscious mental experiences. When Steve Link questioned me on the matter, I found I had a 'gap in knowledge' when it came to knowing exactly what Herbart meant when he applied the term "strength" to a mental representation. So I read the untranslated opening of Psychology as Science that preceded Herbart's presentation of his mathematical model. I discovered that the "strength" of a mental representation, A, is the degree to which A resists an opposing force generated by another mental representation, B, that happens to be in consciousness at the same time as A.

Second, also concerning the beginning of the 19th century, the task of psychophysics is to provide a relationship between the (physical) intensity of a sensory stimulus and the (psychological) magnitude of the resulting sensation. Most psychophysicists are probably aware of the distinction between "extensive" and "intensive" measurement-units. Typical extensive measurement-units apply to distance (e.g. millimeters), time (e.g. seconds), and mass (e.g. grams). Typical intensive measurement-units apply to temperature (e.g. degrees Celsius).

But I was not aware, until I read Heidelberger's (1993/2004) biography of Fechner, that it was Immanuel Kant (1724-1804) who first made this distinction. He did so in the first edition (1781) of his Critique of Pure Reason, and reworded it slightly in the second edition (Kant, 1787/1929).

I also discovered that I had a 'gap in knowledge' concerning the contributions of William Whewell (1794-1866) to our understanding of the importance of measurement in the history of science generally (Whewell, 1847/1967). He stressed that extensive magnitudes are numerical measurement-units (e.g. millimeters), whereas intensive magnitudes are best represented in terms of "degrees" (e.g. degrees Celsius).

Add a rod that is one unit in length end-to-end to another rod of the same length. The two rods have a total length of two units. Add water at 10 degrees Celsius to an equal volume of water at 10 degrees Celsius. The volume (which is an extensive measurement-unit) doubles. But the temperature (which is an intensive measurement-unit) does not. The water's temperature remains at 10 degrees Celsius.

Third, concerning Fechner's work in the middle of the 19th century, I wondered whether psychophysics had deeper roots in the academic discoveries of physicists than I had realized. The law that is currently called "Weber's Law" by psychophysicists was discovered in experiments carried out by Ernst Heinrich Weber (1795 – 1878). He successively held positions as Professor of Anatomy and Professor of Physiology at the University of Leipzig. He had a younger brother, Wilhelm Weber (1804 – 1891), whose mathematical gifts led him to work with the great Karl Friedrich Gauss (1777 - 1855) at the University of Göttingen. During his long career, Wilhelm Weber studied measurement-units in various branches of physics, including electricity and magnetism. Moreover, there exists a second "Weber's Law" (due to Wilhelm) that concerns the forces exerted on each other by two electrically charged masses in relative motion (Jungnickel & McCormmach, 1986, Vol I, pp 138-143).

At an early stage in his career, however, Wilhelm Weber became involved in a political controversy concerning academic freedom. He resigned from his Göttingen post in 1837 and went to live in Leipzig. In 1843, he was offered the chair of physics at the University of Leipzig. This chair was unexpectedly vacant. Its holder had been none other than Gustav Theodor Fechner (1801 – 1887), who had resigned because of ill health. Fechner had merited that position because of his work on Ohm's Law. My 'gap in knowledge' came about because I had not noticed that E. H. Weber, G. T. Fechner, and Wilhelm Weber all resided, and probably socialized, in Leipzig from 1837 – 1849. In that year, Wilhelm Weber left Leipzig to be reinstated in his old position at Göttingen. Two years later, on October 22, 1851, Fechner had his insight into how physics and psychology might be partnered.

Fourth, concerning the end of the 19th century, following his acceptance of one of the chairs in philosophy at the University of Leipzig in 1875, Wilhelm Wundt (1832 – 1920) founded his famous Institute of Experimental Psychology there in 1879. Here, he supervised a large number of doctoral dissertations, some of which concerned psychophysics.

At about the same time, in 1881, G. E. Müller (1851 – 1934), whose doctoral thesis had been supervised by R. H. Lotze (1817 – 1881), took over Lotze's position as chair of philosophy at the University of Göttingen. Müller, like Wundt, supervised doctoral research on psychophysics as well as on human memory. In one psychophysical study, Müller & Schumann (1889) demonstrated that a participant's judgement, on a particular trial T, of which of two lifted weights felt the heavier, could be biased by the felt heaviness of weights lifted on trials preceding trial T. In another study, Martin & Müller (1899) found that judgements of which was the heavier of two weights were not necessarily based on a mental comparison of the two feelings of heaviness induced by the two weights. The judgements, instead, could be based on a "general impression" of the heaviness of just one of the weights. In both studies, neither of which is available in English, it was clear that paired-comparison judgements were influenced by subjective cognitive factors over and above the objective difference in the actual weights of the two stimuli.

My 'gap in knowledge' in this context was that I did not notice that some of G. E. Müller's graduate students and co-workers might have been motivated by this evidence to achieve loftier goals. For example, in Müller's laboratory, Husserl (1900-1901) wrote a book that started the movement known as phenomenology. Schumann went on to design the tachistoscope used by Wertheimer (1912/1961) in the study of apparent motion that ushered in the Gestalt movement. Other Gestalt psychologists who were trained in Müller's laboratory included David Katz (1884-1953) and Edgar Rubin (1886-1951).

More details about these matters will be found in a forthcoming volume in the Scientific Psychology Series, edited by ISP members Stephen W. Link and James W. Townsend. The book is titled *The Creation of Scientific Psychology* (Murray and contributions by Link, 2021).

Acknowledgement

I am grateful to Rachel Breau, Stephen Link, and Susan Ward for their assistance in preparing this article.

References

Fechner, G. T. (1964). Elemente der Psychophysik [Elements of Psychophysics] (2 vols). Amsterdam, Netherlands: J. Bonset. (Original work published 1960)

Heidelberger, M. (2004). Nature from within: Gustav Theodor Fechner and his psychological worldview (C. Klohr, Trans). Pittsburgh, PA "University of Pittburgh Press. (Original work published 1993)

Herbart, J. F. (1890). Psychologie als Wissenschaft [Psychology as Science]. In K. Kehrbach & O. Flügel (Eds.), Jon. Fr. Herbart's sämtliche Werke. In chronologischer Reihenfolge. (Part 1. Vol. 5, pp. 177 – 434). Langensalza, Germany: Hermann Beyer und Söhne. (Original work published 1824)

Jungnickel C. & McCormmach, R. (1986). The intellectual mastery of nature: Theoretical physics from Ohm to Einstein. Vol 1. The torch of mathematics 1800 – 1870. Chicago, IL : Chicago University Press.

Kant, I. (1929). Critique of pure reason (N. K. Smith, Trans.). London, UK: Palgrave Macmillan. (Original work published 1781; 2nd ed., 1787).

Martin, L. J. & Müller, G. E. (1899). Zur Analyse der Unterschiedsempfindlichkeit: Experimentelle Beiträge [Towards the analysis of sensitivity to difference: Experimental contributions]. Leipzig, Germany: J. A. Barth.

Müller, G. E. & Schumann, F. (1889). Ueber die psychologischen Grundlagen der Vergleichung gehobener Gewichte [On the psychological foundations of comparisons between lifted weights]. Archiv für die gesammte Physiologie der Menschen und der Tiere, 45, 32 – 112.

Murray, D. J. and contributions by S. W. Link (2021). The creation of scientific psychology. Abingdon, UK : Routledge (Taylor and Francis).

Wertheimer, M. (1961). Experimental studies on the seeing of motion. In T. Shipley (Ed., Trans., abridged): Classics in psychology (pp 1032 – 1084). New York, NY: Philosophical Library (Original work published 1912).

Whewell, W. (1962). The philosophy of the inductive sciences founded upon their history (2 vols.) (2nd ed.) New York, N.Y: Johnson Reprint Corporation (Original work published 1847)

Black Boxes of the Mind: From Psychophysics to Explainable Artificial Intelligence

Jordan Richard Schoenherr [1,2]
[1]Department of Psychology / Institute for Data Science, Carleton University
[2]Army Cyber Institute / Behavioral Science and Leadership Department. US Military Academy
Jordan.Schoenherr@carleton.ca

Abstract— For more than a century, the approach embodied by psychophysics have attempted to identify a mathematical function that can both describe and predict the relationship between the input and output of a system. Contemporary algorithmic approaches in data science such as machine learning and neural nets ask many of the same questions while ignoring the lessons from psychophysics. In this review, I identify parallels between the 'black box' algorithms and drive toward explainable artificial intelligence (XAI; Vilone, & Longo, 2020) with early approaches to psychophysics and neuroscience and how these mathematical representations were interpreted. Using these historical discussions as a point of reference, I discuss the constraints that need to be placed on explainability.

Keywords— Explainable AI, psychophysics, black box algorithms, levels of processing

I. INTRODUCTION

Psychophysics represents one of the first attempts to systematically account for the relationship between input and output in our perception of the world. Both Weber and Fechner sought to establish a generalizable mathematical relationship between stimulus and response. However, despite the persistence of psychophysical methods, psychophysics is a relatively small, heterogeneous discipline (p. 292, Teghtsoonian & Teghtsoonian, 1989; see also, Scheerer, 1987). In sharp contrast, contemporary work in the areas of data science and machine learning represent a growing field that is considered to have ever-increasing importance. Although the similarities between these fields of research should not be overemphasized, machine learning, artificial intelligence, and neural networks (collectively, autonomous/intelligent systems, or A/IS)[1] share the central goal of psychophysics: to establish equation(s) that accurately described and predict the relationship between inputs and outputs.

In the case of A/IS, recent concerns have been expressed regarding the extent to which developers or users truly *understand* the properties and operations of resulting networks. According to recent commentators, A/IS merely identify patterns and fail to connect these patterns to prior knowledge structures (e.g., Leetaru, 2019). In more concrete terms, although the operations and products of machine learning algorithms or neural networks can be *interpreted*, this does not mean they provide an adequate *explanation*. Moreover, even if an explanation might be *believable* within a community of practitioners or by the general public, this in no way implies that it provides an accurate or valid explanation. The distinction between interpretability, explainability, and believability prove crucial when attempting to understand what we are trying to explain with a mathematical representation. Despite the importance of explainable for humans, the extent to which A/IS need to be explainable requires a consideration of the context in which these systems are used (e.g., pattern recognition, financial decisions, medical diagnoses), reflecting more general concerns of scientific explanation (e.g., Salmon, 1989/2006).

The debate over the importance of interpretability and explainability have many precedents in psychophysical research. Early debates in psychological science concerning the adequacy of psychophysical approaches and 'black box' learning mechanisms of North American behaviorism can inform the development of explainable artificial intelligence (XAI; Došilović, et al., 2018). In this review, I consider three strands of research: psychophysical attempts at describing and explaining mental processes (e.g., Fechner, 1860), neuroscientific considerations of levels of processing (e.g., Marr, 1982), and current approaches to machine learning (ML), neural nets (NN), and artificial intelligence (AI). Specifically, despite the apparent belief that "data science" reflects a comparatively new field, I will argue that both it and the recent movement toward XAI reflect a continuation – albeit, a much different – of the concerns of psychophysics (e.g., classification, pattern identification, discrimination).

II. PSYCHOPHYSICAL APPROACHES TO MENTAL PROCESSES AND REPRESENTATIONS

Despite being a "transdisciplinary research program," (p. 1211, Ehrenstein & Ehrenstein, 1999), psychophysics concerns itself with the identification of lawful mathematical relationships between intensity of the external stimulus to the magnitude of subjective response. In contrast to the historical oddity of behaviorism that denied the importance of understanding mental operations (Skinner, 1976; Watson, 1913; cf. Hull, 1929), Weber and Fechner believed that equations could capture the operations of sensory processes. Weber suggested that a ratio could be identified that described the changes in objective intensity of an environmental stimulus to the subjective perception of the changes detected by the individual. However, it would only be through Fechner's work that this equation was formalized (see below). However, debate continues in terms of whether a universal law can be identified (e.g., see Krueger, 1989). Thus, rather than a 'black box', Fechner wanted to develop an equation that *described* the mind.

[1] Autonomy and intelligence represent two, distinct of a system. Autonomy represents the extent to which a user needs to intervene in the operations of a system, i.e., a human-in-the-loop. Intelligence reflects the extent to which a system can effectively perform a task, i.e., a systems accuracy, speed relative to a referent system (e.g., humans).

Crucially, although one could interpret psychophysics to reflect a mathematical approach that is otherwise fundamentally equivalent to behaviorism, it is important to understand the context of this research. Fechner's psychophysical work was concerned with low-level perceptual phenomena, not higher-order knowledge structures and multiple cognitive processes (cf. Fechner, 1851). Discounting Fechner's work on these grounds is to impose over a century of the cognitive paradigm to a pre-paradigmatic era of experimental psychology that drew on numerous theories and methods (Ehrenstein & Ehrenstein, 1999). Unlike the behaviorists, he was expressly concerned with internal mental operations, i.e., 'inner psychophysics'. As Heidelberger (2004) notes "[f]rom Fechner's standpoint, one must first work out a great amount of preparatory theory before setting up a [psychophysical] scale … a central task of psychophysics is to develop a generalized form the measurement of structures of difference; particularly formulated in terms of probability," (p.205).

A comparison of his work with that of Ernst Mach provides a useful illustration of this. Adopting a physical approach, Mach also sought a unification between physical and psychological laws, attributing his interest to Fechner's *Elements* (Mach, 1984). His approach was less concerned about mental states than Fechner. Indeed, Marr (2003) notes that American behaviorism was influenced by Mach, often using him as a means to justify a 'descriptive approach' to mental processes. According to this interpretation, he was interested solely in inductive principles and not the kind of deductive theories typically associated with explanation. Specifically, Mach viewed Explanation as "nothing but condensed description" (quoted in Cohen, 1970, p. 136). Thus, for Mach, to comprehensively describe the phenomena is to provide a sufficient explanation.

A more contemporary framework that considers the objective of explanations in psychophysical research was provided by Krueger (1989). Reviewing evidence to support the existence of a universal psychophysical law, Krueger (1989) notes that: "the true psychophysical function … [accomplishes] three objectives, which relate to the (successively more important) predictive, descriptive, and explanatory aspects of the problem," (p. 251). For Krueger, whereas *prediction* implies a point-to-point mapping, *description* requires that such a mapping can be defined by a mathematical expression, and an *explanation* requires presenting how external stimulus are converted into internal represents and processes ("events" in Krueger's terminology).

Consequently, in contrast to Mach, Krueger's account of the epistemology of psychophysics includes more than simple prediction and description, it also requires an explanation of the conversion process. However, this account is restricted to low-level sensory and perceptual phenomena. Thus, only sensation and perception are contained within the scope of psychophysical explanation. Indeed, even Fechner's consideration of aesthetic judgments (Fechner, 1871, 1876) assumed that they should be derived from "from below" through inductive methods rather than "from above" using philosophical principles. Indeed, the top-down approach that Fechner (1876) rejects could be considered an explanatory method, i.e., explaining why colours, composition, etc. were aesthetically pleasing to an observer. However, the focus on basic sensory phenomena necessarily constrains psychophysical theory, creating trade-offs in terms of the kinds of explanations that can be offered. Namely, scientific explanations are useful to the extent that they provide missing information (e.g., Salmon, 1989). Consequently, if psychophysical theory is focused on description and predictions in terms of mathematical representation, it is undetermined (Quine, 1975) such that multiple theories (i.e., neuroscientific, cognitive) can be used to explain the phenomena.

III. Neuroscience and Cognition

Beyond psychophysics, others have considered and attempted to integrate the multiple level of analysis that can be used to understand the relationship between input-output. The most prominent contemporary account of kinds of explanations in sensory and perceptual research was provided by Marr (1982). Marr claimed that there were three levels of analysis: computational, algorithmic, and implementational. At the *computational level*, a problem is described in a general (abstract) manner, i.e. an equation. For Marr, this represents "the most abstract is the level of *what* the device does and *why,*" (p. 22). At the *algorithmic level*, a problem is described in terms of the processes required to turn an input representation into an output representation. Finally, at the *implementational level*, a problem is described in terms of the substrate (e.g., neurons, neuroanatomical structures) that represent the *physical* realization of computational and algorithmic levels. For Marr, "Vision is . . . first and foremost, an information-processing task." (p. 3). At a basic level, this means that an environmental stimulus reflects an input whereas a participant's response reflects an output, e.g., "A computational analysis will identify the information with which the cognitive system has to begin (the input to that system) and the information with which it needs to end up (the output from that system)," (Bermúdez, 2005, p. 18). Consequently, traditional psychophysical explanations are straightforwardly located at the computational level of analysis leaving the corresponding cognitive and many perceptual processes (e.g., algorithmic level) and their physical architecture (e.g., implementation level) unaddressed.

A straightforward means to bypass the discontinuities between these levels of psychological explanation is Pylyshyn's (1984) notion of cognitive impenetrability. For Pylyshyn, cognitive architecture is "the basic operations for storing and retrieving *symbols*," (p. 30; emphasis added). In contrast, sensory signals, and some perceptual processes ('functional architecture'), are typically non-symbolic, i.e., a pattern of activation. Discrepancies will arise between a cognitive systems symbolic interpretation of sensory and perceptual processes which are non-symbolic. For instance, subjective reports of response confidence demonstrate miscalibration, i.e., a participant's confidence ratings (e.g., 50%, 60%, 70%) does not correspond to their accuracy which can be attributed to failures of mapping confidence categories onto subjective experience (Schoenherr, 2019). This suggest that sensory and most perceptual processes cannot be directly accessed by cognitive and metacognitive systems.[2] Instead, a translation or rescaling process is required (Schoenherr & Petrusic, 2015). Along these lines, researchers have considered conscious experience as a result of a progressive aggregation of signals between subsystems. In each sensory, perceptual, and cognitive subsystem, there are 'microconsciousness' as information is integrated together and made available to higher-order mental processes (Daehane et al., 2006; Zeki, 2003; see also, Dennett, 1991). Consequently, any description or report of the mental states or operations reflects an *indirect measure* of the representations or processes concatenated from early levels of processing (Tunny & Shanks, 2003).

Adopting a computational level of analysis can additionally assist in reducing the appearance of incommensurability between theories of sensation and perception. Consider models of categorization. At the sensory and perceptual level, categorization requires the retention of a prototypes, rules, or instances in memory and that these representations are compared via a decision-making process following the discrimination or identification of a stimulus presented in the immediate environment (for a review, see Goldstone et al., 2012). Although many researchers focus on the possible different representations, there have also been demonstrations that, at a (mathematical) function level, these representations are related (Ashby & Maddox, 1993). Although these categorization models consider are comparatively simple relative to contemporary models of categorization (e.g., SPEED; Ashby et al., 1998; Ashby et al., 2007), such attempts following earlier attempts at providing a mathematical unifying framework (Krueger, 1989). Thus, although psychophysical research confines itself to a computational level concerned with descriptions and predictions, it can be mapped onto other levels of analysis.

IV. Autonomous and Intelligent Systems

Starting with Charles Babbage's conception of the Difference Engine and Analytical Engine and following Alan Turing's more successful realization of these systems, computer scientists have sought to develop computing machines that, when presented with input, can provide the user with a solution. Simplifying the numerous available approaches, I will consider Machine learning (ML), artificial intelligence (AI), and neural nets (NN) reflect three kinds of autonomous / intelligent systems (A/IS). Despite the potential for an overriding connection between these kinds of systems (Domingos, 2015), these and other approaches to [problem-solving] have a number of distinct features that I will consider in terms of their explainability.

Artificial Intelligence. McCarthy et al. (1955) introduced the idea of 'artificial intelligence' suggesting that:

> "every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."

Arguably, the focus on the definition is on *human-like* symbolic intelligence (i.e., language, abstractions, and concepts). Thus, AI are inspired by human and nonhuman animal sensory, perceptual, and cognitive processes. They often have dedicated processes such as memory, attention, executive function, etc. (e.g., Anderson, 1996; Laird & Newell, 1987). The products of dedicated subsystem (or modules) are typical used in some higher-order operation analogous to executive function and response selection. Due to this level of analysis, AI does not share many features with psychophysical models, i.e., they are typically considered in terms of *cognitive* similarity or plausibility. (cf. hybrid models that include associative learning; e.g., Thomson & Lebiere, 2013). Consequently, these systems are both highly interpretable and explainable in that they (correctly or incorrectly) link to existing theory about cognitive processes (e.g., attention, memory).

Neural Networks. Although contemporary deep NN approaches have overshadowed it, their precursor such as the Perceptron, connectionist networks, and parallel-distributed processing have many of the same issues (see also the Pandemonium Model; Selfridge, 1959). These approaches assume non-symbolic processing wherein knowledge is stored and processed in terms of units that have connections which are assigned different weights between an input and output layer. For instance, the connections of the Perceptron (Rosenblatt, 1958; Figure 1) provide a physical instantiation of this idea. In that these systems can be directly examined, the patterns of activation can be known, connections between units and layers can be identified, and the output can be observed. Although relatively limited, more generalized forms of learning that introduce multiple processing layers (i.e., hidden units), parallel distributed processing, and activation and inhibition extend this approach and more closely approximate the global function of neurons and gross neuroanatomy such as parallel distributed processing (PDP; McClelland & Rumelhart, 1986). In line with Marr's account, such models present *implementation-level* explanations.

Unlike AI, NN are not necessarily developed to reflect human cognitive architecture. However, units and layers can perform comparable functions to cognitive operations (e.g., attention, identification, response selection). Thus, NN and the deep learning variants can be explained using neuron- and neuroanatomical-like metaphors. However, this level of analysis (Marr, 1982) might not reflect the kind of explanation that is useful. Thus, although their operations can be described mathematically, they are not analogous to the early attempts at psychophysics.

Machine Learning. In contrast to Ai and NN, ML reflect systems that are the most removed from the structure and function of biological systems (e.g., expert systems). A discussion of the variety of ML approaches is outside the scope of the present paper. It is nevertheless useful to note that many approaches to ML often reflect repackaging of existing statistical techniques, e.g., regression, multidimensional scaling, [cluster analysis], Bayesian. For instance, the strength of the Bayesian approach is often claimed to be its generality, e.g. "[t]he problems of core interest in other areas of cognitive

[2] There is no hard distinction between sensation, perception, cognition, and metacognitive. Like others (e.g., Dahaene et al.) I assume that there is an aggregation of signals by subsystems that can be more or less observed across levels of mental processes.

science may seem very different from the problem of color constancy in vision, and they are different in important ways, but they are also deeply similar," (p. 2, Griffiths et al., 2008).

In line with Occam's Razor, it is the parsimonious nature of these models that make them both elegant and robust. Thus, while there are strengths and weakness associated with each (Domingos, 2015), in many cases they are simple enough to ensure a reasonable degree of transparency. However, this does not necessitate that they provide an adequate explanation of the phenomena that they claim to describe. Moreover, simplicity can often obscure numerous assumptions (Jones & Love, 2011). It is not perhaps surprising then that these approaches and concerns find precedent in early psychophysics. As Murray (1993) notes[3], in Fechner's later work (Fechner, 1882), there was evidence that anticipated connectionist models of learning and memory.

V. Explainable Artificial Intelligence

In order to counter the growing complexity and opacity of machine learning systems, researchers and organizations have turned their attention to making these systems *explainable*, i.e., a degree of transparency in terms of understanding the function and operations of these systems. For instance, if an image is misclassified as an instance of 'Category A' rather than 'Category B', the basis for this failure might not be understood. Taking AI to be a general term to subsume these systems, this approach is referred to as explainable AI (XAI; Gunning, 2016). Despite the recent interest in XAI, this field does not represent a new area of inquiry, with precedents in early expert system development (Goebel et al., 2018). For instance, in the context of medicine, understanding how a decision support system arrives at a conclusion was deemed to be a critical feature of their design in order to promote transparency and trust (Gorry, 1973).

What constitutes explainability remains an open question with machine learning papers often providing explanations that are transparent to their own community, e.g. detection of a feature at Location X results in an increased probability of classify an image as a member of Category A. Here, we can make a useful distinction between interpretability and explainability (Schoenherr & Thomson, 2020). Interpretability parallels Krueger's (1989) notions of prediction and description. Specifically, an individual's ability to interpret a system requires that they can identify a lawful relationship between input and output and capture this relationship in an equation that can make predictions.

In that psychophysicists, neuroscientists, or A/IS developers can identify psychophysical functions, neuroanatomical structures, and algorithms that allow them to make predictions, they certainly can be said to understand properties of the respective systems. However, like Krueger, this understanding does not imply that an adequate explanation has been provided. The input-output mapping, however elegant in its mathematical form, scarcely exceeds the explanations of a perceptron. As Teghtsoonian (1974) notes concerning matching functions "Whatever theory one may entertain… it is clear that they describe only the relation between physical events." This observation is crucial in terms of understanding the kinds of explanations that those in A/IS can provide. Like the 'hard problem' of consciousness (Chalmers 1995), XAI must consider the potential gap between what is intelligible and descriptions that are accurate representations of the system's operations.

Interpretability can be understood in terms of correlation and regression. A correlation describes a relationship between two variables but does not illustrate deep causal relationships. Similarly, a regression equation is used to predict Y i given Xi but does not explain it. Consider a general, multiple regression equation (f(Y)). The parameters of the equation (β_0, $\beta_1 X_1$, $\beta_i X_i$, and ε) are often thought to express a causal relationship. However, this is causation in a relatively weak sense in the absence of linking each variable (and weight) to a corresponding physical or psychological phenomenon. Thus, it is simply a description of the phenomenon which allows for prediction, i.e., an explanandum not the explanans. If we assume that each variable and its corresponding weight represents a given sensory, perceptual, or cognitive process, we are then required to unpack *that* relationship further. Fechner's law can be viewed in a similar manner:

$$p = k \ln\frac{S}{S0}$$

Namely, the subjective sensation (p) is given as the log of stimulus intensity and a constant (k) defined for a given modality is simply the relationship between intensity and perception. This is a predictive description, not an explanation.

In contrast to interpretability, explanations can take a number of forms. For instance, Dennett (1987) considers three kinds of 'stances' that we can adopt to understand a phenomenon: mechanistic (causal relationships), functional (systems' operations), and intentional (mental states). In that psychophysical phenomena are comparatively brief events, such that a detection threshold is often below the subjective threshold for awareness, intentional explanations are not terribly relevant to psychophysics. As I've noted above, psychophysical explanations are also not especially concerned with functional systems. Thus, conforming to Gescheider's (1997) observation, if psychophysics can be said to provide explanations and not simply describe and predict phenomenon, its explanations are local, mechanistic explanations.[4] However, like Bayesian models, in the absence of attempts to align the mathematical frameworks with existing cognitive and neuroscientific models, it is not clear that such mathematical models provide more than a description and predictions (Krueger, 1989; Teghtsoonian & Teghtsoonian, 1989).

Goals of XAI, Lessons from Psychophysics. The issues faced by psychophysical explanations parallel those of XAI. When examining a deep NN or ML systems, the kinds of explanation that is possible is relative. Explanations for NN, or deep NN, can use analogies related to the function of neurons, cell assemblies, and neuroanatomy. Thus, a variety of mechanistic explanations and functional explanations can be provided. Failures in learning, response selection, etc. can therefore be understood in a comparable manner. In contrast, like psychophysical explanations, explainable ML will likely face

[3] It should be noted that Murray only mentions this in passing and directly states that this was not a central concern.

[4] Local in the sense that they are specific to a modality or stimulus dimension.

many challenges. Local causal explanations will likely be plausible, i.e., an error occurred because Process X failed to initialize. However, much like the notion of cognitive impenetrability (Pylyshyn, 1984), it is not clear that the algorithmic processes of ML will be intelligible to humans. Namely, the process of sensory transduction can be described to an observer but this does not explain the subjective experience of vision to the observer in a manner that they can understand and act on. Similarly, if an A/IS makes decisions, to what extent does it need to be transparent in a way that sensation and perception cannot? At present, A/IS might best be seen as extensions of our sensory and perceptual systems, i.e. a form of extended perception and cognition. Consequently, in a very real sense, we might need to accept the impenetrability of these systems.

It could be argued that A/IS are involved in more consequential decisions (e.g., finance, policing, auditing, driving) that can result in a loss of life or livelihood. However, it is important to note that the phenomenon that psychophysics studies are equally consequential. Line lengths discrimination, absolute identification, and stimulus detection provide the basis for consequential decisions such as identifying a light as red or the sound of a horn warning us to apply pressure to a break in the car we are driving. Thus, unlike sensory and perceptual processes, the imperative of XAI likely stems from the lack of familiarity of these systems and a correspondingly low level of trust.

In this way, XAI might be seen as analogous to metacognition. First, we accept the impenetrability of an A/IS based on a NN or ML architecture. This does not mean that these systems can be known in whole (e.g., AI) or in part (e.g., NN). However, it suggests that new criteria should be identified. Transparency and explainability of the underlying operations need to be relative to the system being considered. For instance, informing someone that rods take longer to activate than cones making dark adaptation more problematic around dusk and dawn, can be used to create compensatory strategies.

Second, we must also be cognizant of whether simplified explanations are desirable. For instance, Teghtsoonian and Teghtsoonian note that there is "a dark side" of oversimplification can be equally problematic in XAI. If individuals that are developing or using these systems provide "just so stories" about their operations, such explanations might be believable and adequate in a given context but might fail to generalize, i.e., they are not universalizable. This is crucial consideration given the problem of scalability: What might work on one level with a given set of data might fail at another.

VI. Conclusion

Understanding the history of psychophysics, sensation, and perception is not only important for understand the scope of research within this research tradition, it can also provide insight into the nature and limits of scientific explanation. In the light of this review, the contemporary questions and concerns associate with A/IS appear to be nothing new. At the core of approaches such as AI, ML, and NN is the goal of understand the relationship between input (i.e., stimuli, data sets) and output (i.e., binary decision, classifications, behavioural responses) in terms of a mathematical model. Each one of these approaches varies in terms of its resemblance to the psychophysical approach with ML most resembling psychophysics in terms of a focus on providing mathematical solutions to problems and placing less emphasis on identifying and understanding the underlying systems. Similarly, AI bares the least resemblance to the psychophysical approach in terms of its emphasis on functional systems that approach human and nonhuman animal cognitive processes.

Scientific Explanation in Data Science. My reading of the philosophy of science that informs psychophysics and A/IS is that they both stem from the same fundamental questions: what makes a good explanation and whether it can, or has to be, more than an interpretation restricted to a given academic community.

First, we must acknowledge the inherent limits of the kinds of explanations that can be provided in general. While the foregoing discussion suggests that many purported explanations in psychophysics and XAI are interpretations, we should ask if this in anyway makes them less relevant. This is ultimately dependent on the *purpose* of the explanation. Discussions within the psychophysical community using comparable paradigms likely do not require linkages to other levels of analyses. In terms of making predictions, we similarly find that what matters most appears to be the outcome, i.e. successful prediction. Yet, this later assertion – one often related to the Bayesian approach – seems to ring hollow. Specifically, even if a prediction is exceptionally accurate, the accuracy can always be improved. Seen in this light, all science reflects satisficing – scientists need to stop experimenting and creating models to present findings. Concurrently, in order to improve a prediction, unless we are relying on a stochastic process, hypotheses informed by theories must be constructed and tested. For this reason, accepting that interpretations are both necessary and sufficient *alone* appears inadequate.

Consequently, any mathematical models such as presented by psychophysics and machine learning, ultimately reflect the equivalent of descriptive and inferential statistics. Thus, psychophysical models simply make descriptions and predictions but do not necessarily provide explanations about *why* they were successful in doing so. They are, intentionally or unintentionally, black boxes of learning and decision-making. This need not be a problem in that it provides the kind of inductive empiricism that provides a durable core, around which auxiliary hypotheses informed by theory (e.g., Lakatos, 1976) and instruments (e.g., Laudan, 1977) can change over time. In this manner, the psychophysical approach represents a Hephaestion Paradigm, i.e., creating the data and tools that are often relegated to the background of contemporary psychological science where theoretical debates are prioritized.

Second, there are practical limits to what kinds of explanations can be provided in psychological science in particular. If the majority of sensory and perceptual processes are cognitively impenetrable (Pylyshn, 1989) due to a successive summation and abstraction of information at higher-levels (Daehane et al., 2006), then the kinds of explanations that are intelligible to those outside of psychophysics are highly constrained. As I've established, areas such as metacognition do not represent the proper subject of psychophysics. Interpretative and predictive models of metacognition can be constructed but they ultimately reflect processing of abstract representations that have been simplified, ignoring the representations that psychophysics is concerned with. Here, we see parallels with A/IS. The majority of A/IS currently in operation are designed to identify patterns in the data, not self-regulate nor communicate the reason for their successes and errors. Consequently, any explanation provided by an A/IS is likely to have limited correspondence to its underlying operations. However, metacognition can surely supplement and complement psychophysical observations. In a similar manner, explanations can perform the same function for mathematical models of the relationship between variables that underpin A/IS. Although they might be underdetermined by data (Quine, 1975), they can facilitate our understanding of phenomena.

References

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist*, 51, 355-365.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, 37, 372-400.

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, 105, 442-481.

Ashby, Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. Psychological review, 105(3), 442. F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, 114, 632-656.

Bermúdez, J. L. (2005). *Philosophy of Psychology: a Contemporary Introduction*. New York: Routledge.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127-181.

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2, 200–219.

Dehaene, S., & Changeux, JP., Naccache, Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences*, 10, 204-211.

Dennett, D. C. (1991) *Consciousness Explained*. Boston: Little, Brown and Company.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

Domingos, P. (2015). *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake Our World*. Basic Books.

Došilović, F. K., Brčić, M., & Hlupić, N. (2018, May). Explainable artificial intelligence: A survey. In 2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO) (pp. 0210-0215). IEEE.

Ehrenstein, W. H. & Ehrenstein, A. (1999). Psychophysical Methods, in U. Windhorst and H. Johansson (eds.) *Modern Techniques in Neuroscience Research.* Berlin: Springer, pp. 1211–1241.

Fechner, G. T. (1851). Z*endavesta, oder über die Dinge des Himmels und des lenseits*. Leipzig: Breitkopf und Härtel.

Fechner, G. T. (1860). *Elemente der Psychophysik*. New York: Holt, Rinehart and Winston.

Fechner, G. T. (1871). Ueber die Aechtheitsfrage der Holbeirischen Madonna: Discussion und Acten. Leipzig: Breitkopf & Härtel.

Fechner, G. T. (1876). Vorschule der Ästhetik. Leipzig, Germany: Breitkopf & Härtel. Website: http://ia600305.us.archive.org/13/items/vorschulederaest12fechuoft/vorschulederaest12fechuoft.pdf.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2012). Concepts and Categorization. In Weiner, I.B, Healey, A.J., & Proctor, R.W. (Eds.) *Handbook of Psychology, Volume 4, Experimental Psychology*, 2nd Edition (pp. 607-630). New York: Wiley.

Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., ... & Holzinger, A. (2018, August). Explainable AI: the new 42?. In International cross-domain conference for machine learning and knowledge extraction (pp. 295-303). Springer, Cham.

Gorry, G. A. (1973). Computer-assisted clinical decision making. *Methods of Information in Medicine*, 12, 45-51.

Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *Cambridge Handbook of Computational Cognitive Modeling* (pp. 59–100). Cambridge, England: Cambridge University Press.

Gunning, D. (2016). DARPA XAI BAA. DARPA, 2016. https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf. Retrieved August, 19, 2020.

Hull, C. L. (1929). A functional interpretation of the conditioned reflex. *Psychological Review*, 36, 498-511.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34, 169-231.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: An architecture for general intelligence. *Artificial Intelligence*, 33, 1-64.

Lakatos, I. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge University Press.

Laudan, L. (1977). *Progress and its Problems*. Berkeley: University of California.

Leetaru, K. (2019). Why machine learning needs semantics not just statistics. *Forbes*. https://www.forbes.com/sites/kalevleetaru/2019/01/15/why-machine-learning-needs-semantics-not-just-statistics/#ba5ab1d77b5c. Retrieved July, 3, 2020.

Mach, E. (1984). *The Analysis of Sensations and the Relation of the Physical to the Psychical*. Trans. by C. M. Williams, La Salle: Open Court, 1984.

Marks, L. E. (1794b). *Sensory Processes: The New Psychophysics*. Academic Press.

Marr, D. (1982). Vision. San Francisco, CA: W.H. Freeman.

McCarthy, J., Minsky, M., Rochester, N., Shannon, C.E., (1955). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*.

Murray, D. J. (1993). A perspective for viewing the history of psychophysics. *Behavioral and Brain Sciences*, 16, 115-137.

Pylyshyn, Z. W. (1984). Computation and cognition. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behavioral and Brain Sciences, 22(3), 341–423.

Quine, W. V. (1975). On Empirically Equivalent Systems of the World. *Erkenntnis*, 9, 313-328.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65, 386.

Salmon, W. C. (1989/2006). Four decades of scientific explanation. University of Pittsburgh press.

Scheerer, E. (1987). The unknown Fechner. *Psychological Research*, 49, 197-202.

Schoenherr, J. R. (2019). Metacognitive assessments of performance: the psychometric properties of confidence scales and confidence models. In *Proceedings of the 35th annual meeting of the International Society for Psychophysics*, Antalya, Turkey.

Schoenherr, J. R. & Petrusic, W. M. (2015). Scaling internal representations of confidence: effects of range, interval and number of response categories. In *Proceedings of the 31st annual meeting of the International Society for Psychophysics*, Québec City, Canada.

Schoenherr, J. R. & Thomson, R. (2020). Knowledge-to-Information Translation Training (KITT): An Adaptive Approach to Explainable Artificial Intelligence. In *International Conference on Human-Computer Interaction*. Springer.

Selfridge, O.G. (1959) Pandemonium: a paradigm for learning. In *Proceedings of the Symposium on Mechanisation of Thought Processes* (Blake, D.V. and Uttley, A.M., eds), pp. 511–529.

Stationary Office

Skinner, B. F. (1976/2011). *About Behaviorism*. New York: Random House.

Teghtsoonian, R. (1974) On facts and theories in psychophysics: Does Ekman's law exist? In: *Sensation and measurement: Papers in honor of S. S. Stevens*, ed. H. R. Moskowitz, B. Scharf & J. C. Stevens. Reidel.

Teghtsoonian, R., & Teghtsoonian, M. (1989). Unified psychophysics: Wouldn't it be loverly… *Behavioral and Brain Sciences*, 12, 292-292

Thomson, R., & Lebiere, C. (2013). Constraining Bayesian inference with cognitive architectures: an updated associative learning mechanism in ACT-R. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. 35, 3539-3544.

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition*, 31, 1060-1071.

Vilone, G., & Longo, L. (2020). *Explainable Artificial Intelligence: a Systematic Review*. arXiv preprint arXiv:2006.00093.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158–177.

Modelling Subjective Confidence with an Associative Memory Network

Jordan Richard Schoenherr
Department of Psychology / Institute for Data Science, Carleton University
Jordan.schoenherr@carleton.ca

Zeki, S. (2003). The disunity of consciousness. *Trends in Cognitive Science*, 7, 214–218.

***Abstract*—** Subjective reports have played a central role in the determination of the correspondence of properties of external stimuli and subjective perception. For instance, models of confidence have suggested that participants simply need to rescale stimulus strength onto a confidence scale or rescale stimulus information onto a confidence scale. Concurrently, researchers studying metamemory (e.g., judgments of learning) instead suggest that specialized monitoring processes account for discrepancies between subjective awareness and performance. In the current study, I consider whether systematic biases in confidence calibration (e.g. overconfidence) are introduced as a result of confusion in recalling properties of the primary decision process. Using an oscillator-based model of associated recall, I demonstrate that miscalibration might reflect failures of memory processes. Moreover, by using a separate set of cognitive processes to model confidence and primary decision processes, confidence processes can be seen as at least partially separable from the primary decision.

***Keywords*—** Oscillator-based associative recall, memory, confidence reports.

I. INTRODUCTION

Subjective confidence reports have been used in a wide variety of fields ranging from assessments of general knowledge, memory, and jury decision-making. Confidence reports display a number of systematic patterns of variation. Although participants tend to overestimate their performance, they tend to be overconfident when making difficult decisions and underconfident with easy decisions, i.e., the Hard-Easy Effect (Lichtenstein, & Fiscnhoff, 1977; cf. Kvidera & Koutstaal, 2008). Still other models of metacognition assume that memory processes are central to subjective assessments of performance. For instance, rather than basing subjective assessments of performance on information accumulated during the primary decision, participants might use processing cues when judging their confidence in their performance (e.g., Koriat & Ma'ayan, 2005). Unfortunately, theoretical and empirical work in confidence processing is so widespread in the psychological literature, broad surveys and integrative frameworks are often lacking (cf. nelson & Narens, 1990; for an alternative account, see Schoenherr, 2019).

In the current study, I examine the possibility that the processes involved in the primary decision and confidence are separate. The model I present assumes that participants accumulate evidence and translate this evidence into a representation of response certainty (e.g., a balance of evidence comprised of evidence favouring the dominant and nondominant response alternatives). The representation of certainty is then encoded into a short-term store and then retrieved resulting in a confidence report. In order to maintain neurological plausibility, this model of confidence assumes a parallel associative network is responsible for these memory processes defined by parameters for attention and distinctiveness (Brown et al., 2000). In this account, the miscalibration observed in confidence reports is the consequence of retrieval errors. Two simulations suggest that this associative network-based model of confidence (CONFIDANT) can simulate many patterns of miscalibration and confidence response time.

II. MODELS OF CONFIDENCE PROCESSING

Numerous models of confidence processing and subjective assessments of performance have been proposed (e.g., Baranki & Petrusic, 1998; Link, 1992; Nelson & Narens, 1990). In order to summarize these approaches, Schoenherr (2019) considers these models in terms of three dimensions: direct-scaling, locus of confidence processes, and the sources of information. First, models can be differentiated in terms of whether they assume a direct scaling of stimulus information from the primary decision or whether they instead assume that this information must be rescaled, or otherwise transformed, into a representation of certainty. For instance, SDT-based accounts assume that the strength of the signal is directly scaled onto a confidence response scale whereas accumulator-based models assume that accumulated evidence must be rescaled into a representation of certainty (Baranski & Petrusic, 1998). Second, confidence models differ in terms of whether confidence reports are produced concurrently with, or following, the primary decision. For instance, SDT-based models assume that confidence reflects an automatic, rescaling of stimulus strength thereby requiring no further processing (cf. Pleskac & Buseymeyer, 2010). Alternative models of confidence instead assume that confidence processing occurs after the completion of the primary decision (e.g., Vickers, 1979) or can be altered depending on the demands on the primary decision (Baranki & Petrusic, 1998). The recognition that confidence reports require a unique set of processing resources, introduces the possibility that it is separable from the primary decision. Thus, the final dimension of models of subjective awareness is defined in terms of whether evidence from the primary decisions is the sole source of information that influences confidence or whether additional, nondiagnostic information can influence judgment. For instance, whereas both SDT- and accumulator-based models of judgment assume that the primary decision is the sole determinant of responses certain (cf. leakage-based accounts; Usher & McClelland, 2001) metamemory studies instead focus on the influence of encoding and retrieval fluency (Koriat & Ma'ayan, 2005).

III. Connectionist Models of Confidence Processing

Connectionist approaches to cognition remain an important class of models for understand learning processes. Standard connectionist models have traditional contained four features: an input layer and output layer, modifiable connection weights and a learning-rule that adjust the connection weights until an acceptable level of error is obtained (e.g., McClellan & Rumelhart, 1986). These models have also been used to understand confidence processing (Merkle & VanZandt, 2006; Vickers & Lee, 2000).

Parallel Adaptive Generalized Accumulator Network. Vickers and Lee (1998) developed an accumulator model in the context of categorical discrimination that Vickers and Lee (2000) adapted to predict confidence responses. Like other cognitive models (Chaiken, 1980), the Parallel, Adaptive, Generalized Accumulator Network (PAGAN) uses confidence processing as an *active* monitoring agent that adjusts the response thresholds of the primary decision. PAGAN assumes that a confidence level is computed for a given trial and is then compared to an ideal level of confidence (i.e., a confidence bias). The difference between target confidence (set by the observer) and obtained confidence generates a quantity of under- and overconfidence that are maintained in two separate accumulators. Once one confidence accumulator reaches a set threshold, the response criteria of the primary decision is altered depending on the magnitude of the difference between the two confidence accumulators.

Despite its robustness, the PAGAN architecture has several issues. First, despite the use of confidence processing in a monitoring function, confidence still reflects a by-product of the primary decision. This stands in oppositions to findings that DRT increases with the requirement of confidence processing and the findings of an alterable locus of confidence more generally (Baranski & Petrusic, 1998). A second issue concerns the model's ability to model confidence reporting. Although Vickers and Lee (2000) report that the model can fairly accurately model confidence responses, the only data that is reported concerns speed-accuracy trade-offs. Data for other variables such as primary decision difficulty or response interference are not modeled. Similarly, those findings that are presented only refer to confidence calibration and not the time taken to report confidence, i.e. confidence response time.

Poisson Race Model. More recently, Merkle and VanZandt (2006) developed a Poisson race model of confidence calibration. Specifically, this model was developed to account for multiple dependent variables (accuracy, confidence, and response time). To achieve this, the Poisson distribution was used to model response times for the accumulation of information on two separate counters. As response time increases, participants begin to obtain information favouring the dominant and alternative responses until a criterion is reached. At this time, a response is selected and confidence is computed in terms of Vickers (1979) balance-of-evidence hypotheses. This requires that the threshold of evidence for the dominant response is divided by the total evidence accumulated for both the dominant and alternative responses (e.g., Vickers & Packer, 1982).

In a series of simulations, Merkle and VanZandt (2006) demonstrated their success in modelling data. The Poisson race model effectively captures the Hard-Easy Effect and generated a moderately accurate match to the RT data used in the model although the response time function for the fitted values was far steeper than the observed data. However, the predictions of the model do not conform to a considerable number of findings both for the Hard-Easy Effect and for CRT (e.g., Baranski & Petrusic, 2001).

First, there are several problems with the authors' general observations about the confidence literature. As noted above, the finding of uniform overconfidence in sensory and perceptual tasks does not hold. Many studies, including those conducted by the author of this paper (Schoenherr et al., 2010; Schoenherr, et al., 2018; Schoenherr et al., 2020), have demonstrated complex patterns of under- and overconfidence for discrimination tasks. Moreover, the primary determinant of the Hard-Easy Effect is the difficulty of the task not the task type *per se*.

A second difficulty with the Poisson race model is its assumption about RT distribution. Again, numerous studies have shown that CRT does not monotonically decrease with increasing confidence level. Instead, guessing and certain responses are generally far more rapid than intermediate levels of confidence. Consequently, the findings of Merkle and VanZandt (2006) are not representative of the confidence literature as a whole. Consequently, another model of confidence process is required. Here, I will consider the possibility of miscalibration arising from failures of short-term memory.

IV. Connectionist Models of Memory

An overriding problem faced by many early memory models (e.g., Conrad, 1965; Wickelgren, 1965) is their ability to autonomously retrieve of stored information. For instance, it was unclear how the memory system achieves sequential retrieval of items from memory. One means to overcome the problem of autonomous retrieval, is periodic reactivation of stored items. In the earliest formulations of cyclical reactivation, Estes (1972) proposed that during the encoding process, items were associated with a control node. The connections between the control node and items are then periodically refreshed with errors resulting from perturbations in the order of reactivation. These perturbations are increased as a result of the density of items stored within an arbitrary time interval. In this way, the distinctiveness of a portion of the signal is reduced with increases in the density of items stored within an interval.

The notion of memory item distinctiveness has been incorporated into several models of serial-order memory, based on both the global (Murdock, 1960) and the local properties of a sequence (Neath et al., 2006). Later models have instead used dynamic context signals that use a competitive process of activation and inhibition to determine which item is retrieved from memory. One such model proposed by Brown et al. (2000) assumes that synchronicity of such a dynamic context signal with incoming information provides an elegant means to model serial-order memory in a neurologically plausible fashion.

Brown et al.'s OSCillator-based model of Associative Recall (OSCAR) assumes that a multi-frequency signal (i.e. a set of oscillators) is continuously active during encoding, storage, and recall. As a sequence of items is presented, each item is associated with a given state of these oscillators, like events are associated with the hour, minute, and second hands on a clock. Recall proceeds once these oscillators have returned to their initial state, and the sequence is restarted. Given that some states of the oscillators are similar, order errors can arise due to confusability, e.g., similar states of the oscillators or rapid presentation of items during encoding decrease recall accuracy. Despite later studies questioning the necessity of a temporal component (Lewandowsky et al., 2006) a key advantage of OSCAR is its neurological plausibility and that it can function in parallel with other cognitive processes.

V. A Confidence Processing Associative Network

If confidence reports are principally determined by evidence from the primary decision, models of confidence processing must account for how primary decision evidence is translated into a confidence report if they fundamentally differ from the primary decision (cf. Pleskac & Buseymeyer, 2010). OSCAR presents one means to examine the relationship between the evidence that is accumulated during the primary decision and subjective confidence calibration. In the present study, I assume that primary evidence is transformed into a representation of the balance-of-evidence (*C*). While participants are engaging in response selection, *C* is retained in a short-term store by means of an association the state of a set of neural oscillators (here, OSCAR). Thus, in addition to rescaling confidence, participants must recall *C*. However, given that multiple states are used to determine confidence during the accrual process, confidence categories are confusable, i.e., 60% confidence can be confused with either 50% or 70%. Confusability is a decreasing function of similarity with the recall states. As a result of failures of recall, participants become miscalibrated, leading to over-/underconfidence bias. The resulting model reflects a CONFidence Associative NeTwork, or CONFIDANT (Figure 1).

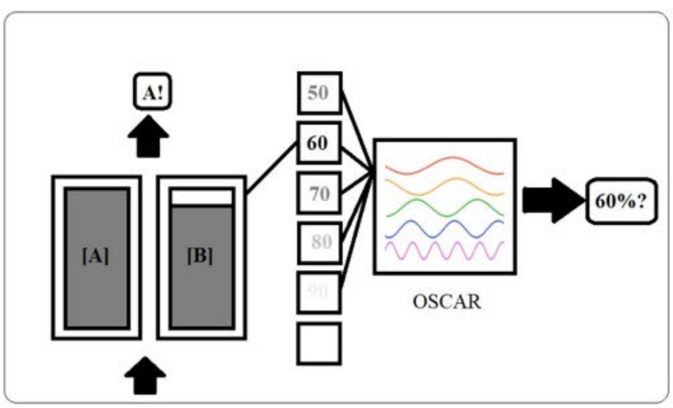

Figure 1. Basic Structure of CONFIDANT. Accumulators produce a response when a response threshold has been reached. Evidence for response alternatives A and B are rescaled onto a confidence scale. Confidence representation is stored temporarily in memory and then recalled.

CONFIDANT. As CONFIDANT is based on OSCAR, the reader is directed to Brown et al (2000) for a greater level of detail. Below I describe the essential components of the model and its implementation. In addition to the typical input and output layers employed by connectionist models, CONFIDANT makes the added assumption that the unique pattern created by the oscillators is associated with an accumulation state to create a learning-context vector.

Like OSCAR, CONFIDANT uses a learning-context signal made up of 15 oscillators that are combined together to create a 16-element learning-context signal. Four out of the fifteen oscillators define each element, with some oscillators contributing more (e.g., θ_1) than others (e.g., θ_8 and θ_{15}). The combinations of a set of these signals constitute the learning-context vector that is associated with the Hebbian learning rule to an information state, *C*. Retrieval depends on reinstating the signal.

Each oscillator is defined by a sinusoid that varies over time defined by $\sin(\phi + t_a{*}\theta_n)$ or $\cos((\phi + t_a{*}\theta_n)$. At the beginning of the simulation, each oscillator is assigned a small random normal value to provide it with an initial state. This initial state is then progressed by adding the product of the distinctiveness parameter (D) and step-size, where step-size (n) is defined as the product of a normally distributed value and 2n. Next, a learning-context matrix, L_a, is created for each time step to which we add the product of the step-size and D to progress it to the nth time step. Having completed the learning-context vectors, a row vector is created for each of the 6 states of the accumulator, i.e., values of *C*.

The nature of the context signal is manipulated by the distinctiveness parameter, *D*, with large values creating a more distinctive signal that permits greater accuracy in recall. Therefore, *D* represents the discriminative difficulty level of a task. An addition source of interference for the association of the context signal with accumulated evidence results from adjusting the attentional parameter (*A*). This parameter changes the rate of decay between the association of the context signal and the accumulated evidence. Reduced attentional resources are modelled by using values of $A < 1$. This describes situations where the primary decision may require additional resources. However, unlike OSCAR, the attentional parameter in CONFIDANT assume values where $A > 1$. This occurs when the primary decision does not require additional attentional resources, such that the confidence process can recruit them during encoding the accumulated evidence. Finally, when confidence is requested, the participant reinitializes the learning-context signal and selects the level of evidence recalled at the time of primary decision response selection.

Differences Between OSCAR and CONFIDANT. Although in many respects CONFIDANT is identical to OSCAR several assumptions differentiate these models. In OSCAR, the weight decay parameter used to model attentional decrement for the *n*th item in a sequence. This constrains OSCAR's attentional parameter between 1.0 (full attention) and 0.0 (no attention). CONFIDANT does not share these constraints. Confidence processing is assumed to constitute an additional operation that is concurrent with the primary decision, comparable to a dual-task paradigm. This allows us to assume that the attentional parameter can vary from little attention ($A < 1$), full attention ($A = 1$), and additional attention that is not being used by the primary decision ($A > 1$). Consequently, when $A > 1$ successive information states reinforce subsequent states.

Another difference between OSCAR and CONFIDANT concerns the assumption about attention. In OSCAR, attention is assigned as a constant value over trials. That is each item in a sequence is assumed to have steadily decrease value (or remain constant in the case were A = 1). However, this reflects an ideal and does not represent actual variations in the attentional resources available. Subsequently, CONFIDANT permits attention to vary around a mean (A*i*) with a standard deviation of 1 (see Experiment 2). All simulation were conducted using MatLab.

VI. SIMULATIONS

Simulation 1: Preliminary Examination of CONFIDANT

Simulation 1 was performed to examine basic properties of CONFIDANT. The primary decision was modelled using an accumulator-based model of decision-making, with two accumulators in a 2-alternative forced choice task (2AFC). Each accumulator had equivalent response thresholds. Following response selection, the balance-of evidence was rescaled onto a 6-point interval scale representing the 6 possible states of a confidence accumulator (i.e., 50% through 100%). This representation was then associated with the oscillator states of the memory model, i.e., OSCAR.

Implementation. In order to simulate the brief presentation of stimuli, the distinctiveness parameter was given a low value (D = 2). Low, intermediate, and high levels of attention were examined by varying the weight decay parameter (A = 0.8, 1.0, 1.2). No response interference was introduced to simulate the presentation of simple stimuli (e.g., lines, squares). The program was executed in a manner described in the introduction and was 15 run. Given that each run represents 6 presentations of information states, this was equivalent to 90 experimental trials.

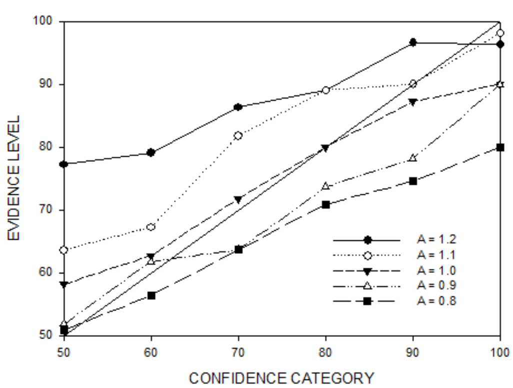

Figure 2. CONFIDANT confidence calibration. Confidence categories retrieved as a function of evidence level provided to the oscillator-based model of memory.

Results and Discussion

Confidence Calibration. CONFIDANT replicated the overall patterns evidenced in studies of calibration (Figure 2). A baseline level of attention (A=1.0) produced nearly perfect calibration with some underconfidence and overconfidence in the start (50%) and end (100%) of the confidence scale, respectively. Altering the attentional parameter resulted in changes in confidence calibration. Increases in the attentional parameter (A*i* = 1.1, 1.2) resulted in greater underconfidence whereas reducing the attentional parameter (A*i* = 0.9, 0.8) increased overconfidence. When attention was comparatively low (A = 0.8), I observed a flattening of the calibration curve. These findings are similar to those observed by Merikle & VanZandt (2006). Contrasted against these are findings where attention is close to an optimal range (A = 1.0).

If CONFIDANT provides a reasonable account of the relationship between the primary decision and confidence processing, this suggests that a key determinant of miscalibration and overconfidence bias is the availability of attentional resources. If participants fail to attend to the diagnostic information favouring either response alternative, greater levels of miscalbration will be observed.

Simulation 2: Modification of CONFIDANT

The results from Simulation 1 are promising, but not without their shortcomings. First, Simulation 1 potentially oversimplifies the relationship between accumulated evidence and its association to confidence level. Presumably, when a stimulus is shown to a participant they wish to accumulate as much evidence as possible. However, they are restricted in the amount of evidence they can obtain due to resource constraints, presentation time, or distractors. Even if the additional assumption is made that participants accrue additional evidence after the primary decision (Pleskac & Buseymeyer, 2010), on any given trial, they will only have observed a sub-set of the total information states available. Thus, it could be argued that if participants only attained evidence for both response alternatives equivalent to a confidence report of 60%, they could not confuse this with 70% given that the information was not provided.

A second issue is that manner in which decay was held constant for each subsequent item, and thus fails to account for natural variability of attention over time. This can be corrected by assuming that attention allocation is described by a normal distribution, where A*i* is the mean attention level used in Simulation 1.

A third difficulty with Simulation 1 is that the rate of accumulation is constant (i.e., D = 2). However, in a comparable manner to attention, difficulty level varies in most experimental and real-world settings. In this way easy decisions will have greater distinctiveness than hard decisions with intermediate decisions lying somewhere in the middle.

Implementation. To simulate the above situation each information state of the balance-of-evidence (50 through 100) was presented individually to CONFIDANT. This means that each simulation included the terminal information state (e.g., 70) and all subsequent values (e.g., 50, 60). After all information states and their confidence levels were generated, the values were randomly presented to the accumulator model and, subsequently, these information states were presented to CONFIDANT on each trial.

Again, I varied attention level (A*i* = 0.8, 1.0, and 1.2) as well as distinctiveness to simulate different difficulty levels (D = 1, 2, 3). The difficulty level parameter was held constant whereas the attention parameter was given by a normal distribution (A*i*, 1).

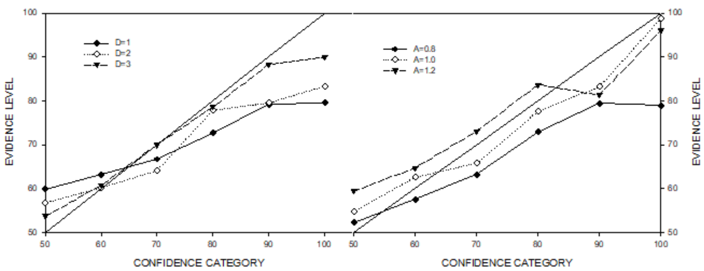

Figure 3. Effects of changes in distinctiveness and attention parameters on confidence response bias.

Results and Discussion

Confidence Calibration. As in Simulation 1, CONFIDANT generated responses that were similar to the patterns observed in experimental data. Replicating Simulation 1, the model's level of confidence varied with attention level. When the attention parameter was increased the simulation generated underconfidence whereas when it was decreased overconfidence was observed. When attention was held at an intermediate level (A*i* = 1.0) the model was reasonably well calibrated. The difference between the results of the current simulation and that of Simulation 1 can be attributed to the introduction of random variation in the attention parameter rather than using a constant value of A*i*.

Crucially, by varying the distinctiveness of the context signal to simulate varying levels of difficulty, Simulation 2 was able to model the Hard-Easy Effect (Lichtenstein & Fischoff, 1977). As Figure 2 demonstrates, high-difficulty items (D = 1) were associated with greater underconfidence relative to low-difficulty items (D = 3).

Table 1. Confidence category CRT (ms) for Simulations 2 and 3.

	Confidence Level					
CRT	**50**	**60**	**70**	**80**	**90**	**100**
Sim 2	922	1131	1115	1271	1127	1410

Confidence Reaction Time. The CRTs provided by CONFIDANT provided a reasonable fit with human data. CONFIDANT effectively captures the CRT patterns for all levels of confidence except those of the 100% confidence category. The present results show larger response latencies for 100% confidence than for any other level of accumulated evidence. In contrast, some studies of confidence have often found that the certainty category is associated with relative smaller response latencies (e.g., Baranski & Petrusic, 2001; cf. Pleskac & Buseymeyer, 2010).

VII. Conclusion

Psychophysical models of confidence are typically based on response strength and evidence accumulation (Baranski & Petrusic, 1998) whereas models of metamemory assume a separate, dissociable confidence process (e.g., Koriat & Ma'ayan, 2005). The purpose of the present study was to develop a model of confidence processing that assumes that 1) the primary decision process provides the primary basis for confidence processing but 2) the requirements of memory and attention can degrade this signal during the process of translation between the primary decision and the representation that is used to inform confidence reports.

CONFIDANT is defined by an accumulator-based decision-making model that can account for choice in a 2AFC task. Once a response has been selected, confidence is computed using a balance-of-evidence (Vicker & Packer, 1982). In the current study, this represents six possible combinations of accumulator states, corresponding to confidence categories between 50% and 100%. Once computed, the confidence representation is encoded into an associative memory and then recalled when confidence reports are solicited.

The results from two simulation suggest that CONFIDANT can account for subjective miscalibration. For instance, patterns of both over- and underconfidence are observed for the extreme values of the confidence response scale, i.e., 100% and 50%, respectively. Moreover, subjective miscalibration also differed depending on the amount of attention and difficulty associated with the task. Finally, unlike many other models of confidence processing (cf. Pleskac & Buseymeyer, 2010), CONFIDANT also provides the basis for confidence response time in terms of the amount of activation. Thus, a primary decision process and an associative memory system might be sufficient to account for the relationship between primary decision and confidence processing. Consequently, metacognition ultimately reflects the relationship between a combination of lower-order sensory and perceptual processes and short-term memory. Consequently, CONFIDANT can be used as a module that can receive input from multiple sensory and perceptual processes.

Future Directions and Concerns. Two more specific concerns can also be raised about CONFIDANT. First, the model assumes that can recall *all* possible accumulator states. However, on some trials, certain accumulator states might not be reached. For instance, if participants obtain 70% confidence, they did not have the possibility of encoding representations greater than 70%, i.e., 80, 90, 100%. Moreover, CONFIDANT also assumes that evidence accumulation occurs in a monotonic manner, i.e., a 50% threshold is followed by a 60% threshold, then a 70% threshold, etc. It is conceivable that there is instead vacillation between states of evidence accumulation. Thus, recall of a confidence representation might be more stochastic and require more accumulation states than are represented here, e.g. 50,6, 50, 60, 70, 60, etc.). However, these assumptions would simply require additional modifications to CONFIDANT.

Second, the current series of simulation was conducted to be parsimonious in terms of the number of assumptions and parameters used to model confidence. It replicates general patterns observed in the confidence report literature. However, it does not model a specific data set. Future empirical studies should investigate the basic assumptions of CONFIDANT in terms of both difficulty, distinctiveness, and attention. Moreover, while I have assumed a balance-of-evidence process here, alternative confidence representations might instead provide a better fit to human data, e.g. using nondiagnostic in doubt-scaling or random-walk diffusion (Baranski & Petrusic, 1998).

A final issue with the current model is a potential lack of parsimony. CONFIDANT assumes that there are two sets of processes that are required for any subjective assessment of performance: a primary decision and a confidence process. Models of confidence have not always made a clear distinction between these processes (Schoenherr, 2019). For instance, Pleskac & Busemeyer (2010) assume that confidence occurs post-decisionally in terms of an extension of the primary decision. In their model, the additional effort in partitioning the distribution of stimulus strength into confidence categories beyond the 2AFC, is left unaddressed as well as evidence that confidence can be computed concurrently with the primary decision. While parsimonious, such models do not address a number of phenomena associated with confidence processing.

Despite these concerns, CONFIDANT provides support for the possibility that confidence represents a secondary process, one that is separable from the primary decision (Chaiken, 1980; Schoenherr, 2019). Consequently, the results of the present study lend support to the proposal that primary decision evidence is not the only information that influences subjective assessments of performance (Koriat & Ma'ayan, 2005). For instance, cues such as encoding and retrieval fluency might be modelled by the attention and distinctiveness parameters used in Brown et al.'s (2000) oscillator-based model of memory.

Acknowledgements

The author would like to express his sincere thanks to Dr. Craig Leth-Steensen for coding the initial version of the OSCAR model for an in-class demonstration.

References

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: 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/Revue canadienne de psychologie expérimentale*, 55, 195.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127.

Chaiken, S. (1980) Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, 39, 752–766.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4, 161-169.

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52, 478-492.

Kvidera, S., & Koutstaal, W. (2008). Confidence and decision type under matched stimulus conditions: Overconfidence in perceptual but not conceptual decisions. *Journal of Behavioral Decision Making*, *21*, 253-281.

Lewandowsky, S., Brown, G. D., Wright, T., & Nimmo, L. M. (2006). Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, 54, 20-38.

Lichtenstein, S., & Fiscnhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, 20, 159-183.

Link. S. W. (1992). *The Wave Theory of Difference and Similarity*. Hillsdale: Erlbaum.

McClelland, J. L., Rumelhart, D. E. & the PDP Research Group (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Vol. 2. Psychological and Biological Models*. MIT Press.

Merkle, E. C., & Van Zandt, T. (2006). An application of the Poisson race model to confidence calibration. *Journal of Experimental Psychology: General*, 135, 391.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review*, 67, 16-31.

Neath, I., Brown, G. D., McCormack, T., Chater, N., & Freeman, R. (2006). Distinctiveness models of memory and absolute identification: Evidence for local, not global, effects. *Quarterly Journal of Experimental Psychology*, 59, 121-135.

Nelson, T., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.) *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 26. Academic Press. pp. 125–169.

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, 117, 864.

Schoenherr, J. R. (2019) Metacognitive assessments of performance: The psychometric properties of confidence scale and confidence models. *Fechner Day 2019*, 71-78.

Schoenherr, J. R., & Lacroix, G. L. (2020). Performance monitoring during categorization with and without prior knowledge: A comparison of confidence calibration indices with the certainty criterion. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*.

Schoenherr, J. R., Leth-Steensen, C., & Petrusic, W. M. (2010). Selective attention and subjective confidence calibration. *Attention, Perception, & Psychophysics*, *72*, 353-368.

Schoenherr, J. R., & Petrusic, W. M. (2015). Scaling internal representations of confidence: effects of range, interval and number of response categories. *Fechner Day 2015*, 36.

Schoenherr, J. R., Waechter, J., & Millington, S. J. (2018). Subjective awareness of ultrasound expertise development: individual experience as a determinant of overconfidence. *Advances in Health Sciences Education*, *23*, 749-765.

Vickers, D. (1979). *Decision Processes in Visual Perception*. New York: Academic Press.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences*, *2*, 169-194.

Vickers, D., & Lee, M. D. (2000). Dynamic models of simple judgments: II. Properties of a self-organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) model for multi-choice tasks. *Nonlinear Dynamics, Psychology, and Life Sciences*, *4*, 1-31.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, 50, 179-197.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review*, 108(3), 550.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology*, 70, 102-108.

Motion Induced Blindness: Luminance Contrast Sensitivity to Increments and Decrements in the Presence of a Motion Mask

Rebecca A. White,[1] Dan Swift,[2] Sofia C. Lombardo,[1] Erika S. Wells,[3] John E. Sparrow,[4] Abby Oostendorp,[1] Joseph A. LaBarre,[1] Andrew J. Kitt,[1] and Wm Wren Stine[1]
[1]Department of Psychology, University of New Hampshire, Durham, NH, USA
[2]Department of Psychology, University of Michigan-Dearborn, Dearborn, MI, USA
[3]Department of Psychological and Brain Sciences, Boston University, Boston, MA, USA 4Social Science Division, University of New Hampshire-Manchester, Manchester, NH, USA
bill.stine@unh.edu

Abstract— That salient visual stimuli vanish in the presence of a motion mask has been described as Motion-Induced Blindness (MIB). We measured the inhibition created by the motion mask as a luminance contrast threshold. Increment targets, presumably stimulating ON-channels, exhibited mask-induced inhibition when in the presence of an increment mask while decrement targets, stimulating OFF-channels, showed no evidence of such inhibition.

Keywords— Motion Induced Blindness, Luminance Contrast Threshold, ON-Channels, OFF-Channels, 4AFC

I. INTRODUCTION

Motion-Induced Blindness (MIB) describes the disappearance of salient stimuli in the presence of a motion mask (Bonneh, Cooperman, & Sagi, 2001). In Bonneh et al.'s (2001) experiments, subjects fixate a point foveally while attending to one or more targets (small yellow dots) in the periphery. A mask of discreet elements (small blue '+', for example) moves coherently such that any element that intersects a target moves behind that target. Typically, within a few seconds of viewing this display, the targets will appear to vanish. In most studies, total disappearance time is measured by asking subjects to depress a response key while the target is invisible.

Grindley and Townsend (1965) first discovered the effect when they presented a target to one eye and a moving mask to the other eye, which they described as movement masking. Ramachandran and Gregory (1991) and Spillmann and Kurtenbach (1992) found that a uniform patch in the periphery would appear to vanish if surrounded by dynamic random-dot noise, with the noise appearing to fill in the patch (see Kawabe & Miura, 2007; Wallis & Arnold, 2008, as well). Ramachandran and Gregory (1991) describe perceptual filling-in as forming an artificial scotoma. Hsu, Yeh, and Kramer (2004; 2006) provided evidence suggesting that perceptual filling-in and MIB shared common mechanisms, which might be expected given that perceived motion would seem to be related to motion energy normalized by flicker energy (i.e., motion contrast: Georgeson & Scott-Samuel, 1999; Rainville, Makous, & Scott-Samuel, 2005; Rainville, Scott-Samuel, & Makous, 2002). Perhaps related to these effects, MacKay (1960) showed that the perception of a retinally-stabilized image of a 10 min of arc diameter wire, which normally would not fade when stabilized, vanished when an unstabilized image was visually scanned.

Disappearance per se would seem to involve both a change in sensitivity to the target's presence and a shift in detection criteria (Caetta, Gorea, & Bonneh, 2007). These changes can be modeled as response gain change, as measured by brightness matching, coupled with a contrast gain change, measured using contrast detection thresholds (Gorea & Caetta, 2009). Bonneh et al. (2001) showed that total disappearance time is positively related to target contrast, negatively related to target size and speed, and positively related to mask contrast, dot density, and speed, as well as exhibiting Gestalt grouping effects for both the target and the mask (see also Graf, Adams, & Lages, 2002, Experiment 2; Mitroff & Scholl, 2005; Shibata, Kawachi, & Gyoba, 2010). Differential effects have been found for both target and mask contrast valence, which should stimulate ON- and OFF-channels (Dolan, & Schiller, 1994; Schiller, 1992; Schiller, Sandell, & Maunsell, 1986; Zaghloul, Boahen, & Demb, 2003), such that decrement masks induce increment target disappearance more quickly than increment masks, and, overall, increment targets disappear more quickly than decrement targets (Stine, Levesque, Lusignan, & Kitt, 2017). Generally, time to initial target fade is inversely related to total disappearance time (e.g., Hsu et al., 2004). Coherently moving masks are less effective than incoherently moving masks (Wells & Leber, 2014; Wells, Leber, & Sparrow, 2011; however, see Hsu et al., 2004; Sparrow, LaBarre, & Merrill, 2017) due, at least in part, to motion adaptation to the coherent mask (LaBarre & Stine, 2019). Similarity between target and mask elements increases target disappearance time (Hsu et al., 2004). Allocating spatial attention to the target increases total disappearance time while removing attention from the entire display decreases the frequency of disappearance but increases the duration of invisibility once the target has vanished (Schölvinck & Rees, 2009). Libedinsky, Savage, and Livingstone (2009) suggest that the effect of the mask is to increase the likelihood of disappearance rather than to decrease the visibility of the target. Interestingly, Dieter, Tadin, and Pearson (2015) demonstrated that the MIB process continues even when continuous flash suppression (Tsuchiya & Koch, 2005) is used to remove the process from visual awareness.

So, contrast valence effects (Stine et al., 2017) on disappearance time are consistent with what one might expect from a simple model where the observer's motion mask response alters contrast gain (cf., Caetta et al., 2007; Gorea & Caetta, 2009). In this experiment we wished to measure contrast gain changes due to the motion mask's inhibition of target visibility as a function of target contrast valence directly. To that end, we created a stimulus with four peripheral dots, or inducers, and a motion mask consisting of 64 incoherently moving dots, all of which were positive contrast (brighter than the background). Four seconds into the trial, the four peripheral inducers were physically removed and, after a variable delay, a single dot, or target, was briefly flashed in one of the original four inducer locations. The target varied in contrast valence, with either positive or negative (darker than the background) contrast. The observer then reported in which location the target appeared, giving a four alternative forced choice paradigm.

Measuring target contrast detection threshold, we anticipated that contrast threshold would be higher following the presentation of inducers than when the inducers were never physically present due to the inhibition of the appearance of the inducers by the motion mask (i.e., due to Motion Induced Blindness). As that inhibition waned, we thought that threshold would decrease with the time interval between the offset of the inducers and the onset of the target. As well, we anticipated that the inducer effect would vary with contrast valence. Finally, replicating previous work, decrements should have a lower threshold than increments (Dolan, & Schiller, 1994; Schiller, 1992; Schiller et al., 1986; Zaghloul, Boahen, & Demb, 2003).

II. METHODS

Participants. Two females and one male, over the age of 18 with normal or corrected to normal vision, participated in the study. The study was approved by the University of New Hampshire Institutional Review Board.

Apparatus and Stimuli. A Dell Dimension E521 computer running Vision Works (Swift, Panish, & Hippensteel, 1997) in Windows XP drove a Mage Systems M21L-H4101 monitor with a 120 Hz refresh rate. The monitor uses a monochrome P46 ultra-short persistence phosphor (yellow-green; CIE x = 0.427, y = 0.543), presenting 800 x 600 pixels with a pixel pitch of 120 dots per inch. Gray scale was rendered using a Vision Research Graphics Gray-Scale Expander VW16 to provide 15-bit linearized depth. A 67 cd/m2 background was continuously present. Positive contrast stimuli (increments) were 120.6 cd/m^2 and negative contrast stimuli (decrements) were 13.4 cd/m^2. The 21" flat-screen monitor was viewed at distance 1 m while using a chin rest.

The mask and inducers were increments while the targets were either increments or decrements. The inducers were four deg of retinal angle from the fixation dot. The mask moved incoherently at 4 deg of retinal angle per s.

Procedure. Sitting in a darkened room with their head stabilized by a chin rest 1 m from the monitor, each participant viewed the central fixation dot and adapted to the background luminance of the screen for five min. Following adaptation, each trial lasted 20 s during which the motion mask was continuously visible while the four inducers were present just during the first four seconds of the trial. At delays of 0.5 s, 3.5 s, 6.5 s, 9.5 s, 12.5 s, and 15.5 s, the target was briefly flashed (250 ms) in one of the original four inducer locations. After each flash an auditory beep indicated to the participant a brief response interval during which he or she could report the location of the target's flash using a keyboard. The participant received feedback after each of the six responses during the trial. The participant was instructed to maintain fixation on the centrally located fixation dot. A 22 s adaptation period followed each trial.

As a function of the time interval between the offset of the inducer and the target onset (0.5 s, 3.5 s, 6.5 s, 9.5 s, 12.5 s, 15.5 s), inducer stimulus contrast (100% or 0%), and the target contrast valence (increment or decrement), the contrast of the target was varied following a weighted up-down adaptive psychophysical procedure (Smith, 1961; Kaernbach, 1991) in order to converge onto a 0.625 probability of a hit. Inducer stimulus contrast (100% or 0%) and the target contrast valence (increment or decrement) were randomly varied across trials while the inducer offset to target onset varied within trials, as mentioned previously.

III. RESULTS

We conducted a probit analysis (Bliss, 1934; Finney, 1971), weighted by the number of presentations for each condition, of the probability of a hit in our four-alternative forced choice paradigm as a function of five independent variables: the log Weber contrast of the target, the time interval between the offset of the inducer and the target onset, inducer stimulus contrast (100% or 0%), target contrast valence (increment or decrement), and participant. The initial full model included the first four independent variables with all of their interactions plus the participants, giving 16 statistical tests. To keep the familywise type I error rate equal to 0.05, we chose a test type I error rate of $0.0032 = 1 - (1 - 0.05)^{1/16}$, using the Šidák (1967) inequality (see Kirk, 2013, p. 183). Using a backwards elimination procedure, the model was fit to the data, the Akaika Information Criterion (AIC) was calculated (Akaika, 1973; 1974), the effect with the largest p value was removed from the model, the resulting model was again fit to the data, and the process continued until all of the remaining effects were significant. This procedure minimized the AIC, which decreased from 645.4 to 637.1.

The presence of the inducers interfered with detecting the increment targets but not the decrement targets (z = - 3.82, p = 9.94 x 10-5), resulting in higher contrasts required for detection (z = 5.28, $p = 1.33 \times 10^{-7}$). Negative contrast targets were easier to detect overall than positive contrast targets (z = 7.64, $p = 2.20 \times 10^{-14}$) and, of course, target contrast was positively related to detectability (z = - 3.82, $p = 8.58 \times 10^{-18}$). The final model gave a Lave - Efron pseudo = 0.399 $\hat{R}^2_{LE}$ (Efron, 1978; Lave, 1970). No other effects were significant. In particular, there was no effect of delay between the offset of the inducers and the onset of the target.

Defining threshold contrast as that inducing a 0.625 probability of a hit, the effects of inducer and target valence on threshold contrast are presented in Figure 1, with 100(1 - 0.0032) = 99.7% confidence intervals estimated using a bootstrap with 1000 resamples. Again, the presence of the inducer raises the threshold for increment targets relative to decrement targets, and decrements have lower thresholds than increments. The effects of inducer to target interstimulus interval and inducer and target valence on threshold contrast are presented in Figure 2, with 100(1 - 0.0032) = 99.7% confidence intervals estimated using a bootstrap with 10000 resamples. As stated earlier, inducer to target interstimulus interval has no effect.

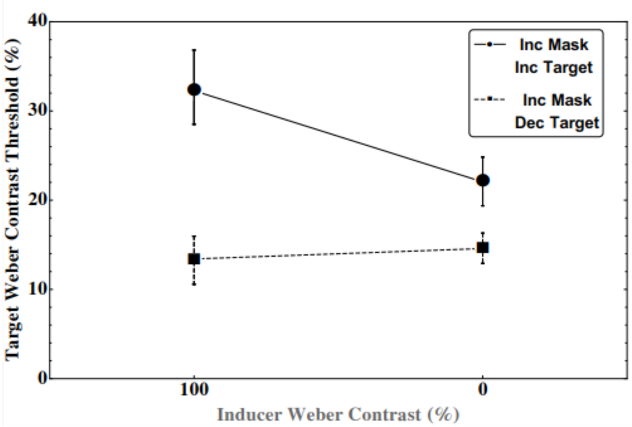

Figure 1. Effects of inducer contrast and target contrast valence on contrast threshold 99.7% bootstrap CI.

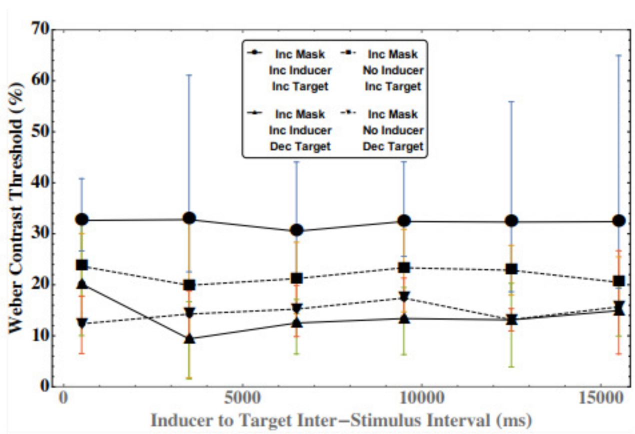

Figure 2. Effects of inducer-target ISI inducer and target contrast threshold 99.7% bootstrap CI.

IV. DISCUSSION

We found that increment targets required greater contrast for detection when an increment inducer had been presented with the increment mask. If no increment inducer was visible, then target increment contrast threshold was reduced. Presumably, residual inhibition of the visibility of the inducers from the motion mask (i.e., motion induced blindness) reduced the target's contrast gain (i.e., raised the target's contrast threshold). This change in contrast gain was evidently long lasting, as there was no effect of the interval between inducer offset and target onset, suggesting that once the inducers have been inhibited, that inhibition can be maintained by the motion mask. Again, without the inducers, threshold was relatively low.

When the target was a decrement, and so presumably stimulated OFF-channels more than ON-channels (Dolan, & Schiller, 1994; Schiller, 1992; Schiller et al., 1986; Zaghloul, Boahen, & Demb, 2003), no evidence of motion induced blindness enhanced inhibition was present. Increment inducers engendered no increase in decrement targets contrast thresholds. As one would expect, decrement thresholds were overall lower than increment thresholds.

These findings are consistent with those of Stine at al. (2017) that the increment – decrement distinction influences motion induced blindness. The technique presented provides a valuable tool for understanding the inhibition induced by the motion mask.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csáki (Eds.), *2nd International Symposium on Information Theory*. Tsahkadsor, Armenia, USSR, September 2-8, 1971 (pp. 267–281). Budapest: Akadémiai Kiadó.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723. doi:10.1109/TAC.1974.1100705

Bliss, C. I. (1934). The method of probits. *Science*, 79, 38–39. doi:10.1126/science.79.2037.38

Bonneh, Y. S., Cooperman, A., & Sagi, D. (2001). Motion-induced blindness in normal observers. *Nature*, 411, 797-801.

Caetta, F., Gorea, A., & Bonneh, Y. S., (2007). Sensory and decisional factors in motion- induced blindness. *Journal of Vision*, 7(7), 1–12. doi:10.1167/7.7.4

Dieter, K. C., Tadin, D., & Pearson, J. (2015). Motion-induced blindness continues outside visual awareness and without attention. *Scientific Reports*, 5, 11841

Dolan, R. P. & Schiller, P. H. (1994). Effects of ON channel blockade with 2-amino-4- phosphonobutyrate (APB) on brightness and contrast perception in monkeys. *Visual Neuroscience*, 11, 23–32.

Efron, B. (1978). Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association*, 73, 113-121.

Finney, D. J. (1971). *Probit analysis. 3rd ed*. Cambridge University Press.

Georgeson, M. A., & Scott-Samuel, N. E. (1999). Motion contrast: A new metric for direction discrimination. *Vision Research*, 39, 4393–4402.

Grindley, G. C. & Townsend, V. (1965). Binocular masking induced by a moving object. *Quarterly Journal of Experimental Psychology*, 17, 97-109. doi:10.1080/17470216508416418

Gorea, A., & Caetta, F., (2009). Adaptation and prolonged inhibition as a main cause of motion-induced blindness. *Journal of Vision*, 9, 1-17. doi:10.1167/9.6.16

Graf, E. W., Adams, W. J., & Lages, M. (2002). Modulating motion-induced blindness with depth ordering and surface completion. *Vision Research*, 42, 2731-2735.

Hsu, L.C., Yeh, S.L., & Kramer, P. (2004). Linking motion-induced blindness to perceptual filling-in. *Vision Research*, 44, 2857-2866.

Hsu, L.C., Yeh, S.L., & Kramer, P. (2006). A common mechanism for perceptual filling-in and motion-induced blindness. *Vision Research*, 46, 1973–1981.

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49, 227-229.

Kawabe, T., & Miura, K. (2007). Subjective disappearance of a target by flickering flankers. *Vision Research*, 47, 913–918.

Kirk, R. E. (2013). Experimental design: Procedures for the social sciences (4th ed.). Washington, DC: Sage.

LaBarre, J. A., & Stine, W. W. (2019). The effects of motion adaptation and contrast polarity on motion-induced blindness. In N. Du Bois, S. Arndt, E. B. Özsoy, S. Bayraktar, E. Gülbetekin, & M. A. Elliott (Eds.). *Fechner day 2019: Proceedings of the 35th annual meeting of the international society for psychophysics* (pp. 44-51). Antalya, Turkey: International Society for Psychophysics.

Lave, C. A. (1970). The demand for urban mass transportation. *The Review of Economics and Statistics*, 37, 320-323.
Libedinsky, C., Savage, T., & Livingstone, M. (2009). Perceptual and physiological evidence for a role for early visual areas in motion-induced blindness. *Journal of Vision*, 9, 1–10.

MacKay, D. M. (1960). Monocular "rivalry" between stabilized and unstabilized retinal images. *Nature*, 185, 834.

Mitroff, S. R., & Scholl, B. J. (2005). Forming and updating object representations without awareness: Evidence from motion-induced blindness. *Vision Research*, 45, 961–967.

Rainville, S. J. M., Makous, W. L., & Scott-Samuel, N. E. (2005). Opponent-motion mechanisms are self-normalizing. *Vision Research*, 45, 1115–1127.

Rainville, S. J. M., Scott-Samuel, N. E., & Makous, W. L. (2002). The spatial properties of opponent-motion normalization. *Vision Research*, 42, 1727–1738.

Ramachandran, V. S., & Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350, 699-702.

Schiller, P. H. (1992). The ON and OFF channels of the visual system. *Trends in Neurosciences*, 15, 86-92.

Schiller, P. H., Sandell, J. H., & Maunsell, J. H. R. (1986). Functions of the ON and OFF channels of the visual system. *Nature*, 322, 824-825.

Schölvinck, M. L., & Rees, G. (2009). Attentional influences on the dynamics of motion-induced blindness. *Journal of Vision*, 9, 1-12.

Shibata, M., Kawachi, Y., & Gyoba, J. (2010). Combined effects of perceptual grouping cues on object representation: Evidence from motion-induced blindness. *Attention, Perception, & Psychophysics*, 72, 387–397.

Šidák, Z. K. (1967). Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626–633.

Smith, K. J. E. (1961). Stimulus programming in psychophysics. *Psychometrika*, 26, 27-33.

Sparrow, J. E., LaBarre, J. A., & Merrill, B. S. (2017). Individual differences in motion-induced blindness: The effects of mask coherence and depth ordering. *Vision Research*, 141, 117–126.

Spillmann, L., & Kurtenbach, A. (1992). Dynamic noise backgrounds facilitate target fading. *Vision Research*, 32, 1941-1946.

Stine, W. W., Levesque, P. A., Lusignan, M. E., & Kitt, A. J. (2017). Motion induced blindness using increments and decrements of luminance. *Proceedings of the Latvian Academy of Sciences. Section B. Natural, Exact, and Applied Sciences*, 71, 372–379.

Swift, D., Panish, S., & Hippensteel, B. (1997). The use of VisionWorks in visual psychophysical research. *Spatial Vision*, 10, 471-477.

Tsuchiya, N. & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature*, 8, 1096–1101.

Wallis, T. S. A., & Arnold, D. H. (2008). Motion-induced blindness is not tuned to retinal speed. *Journal of Vision*, 8, 1–7. doi:10.1167/8.2.11.

Wells, E. T., & Leber, A. B. (2014). Motion-induced blindness is influenced by global properties of the moving mask. *Visual Cognition*, 22, 1–16.

Wells, E. T., Leber, A. B., & Sparrow, J. E. (2011). The role of mask coherence in motion-induced blindness. *Perception*, 40, 1503–1518.

Zaghloul, K. A., Boahen, K., and Demb, J. B. (2003). Different circuits for ON and OFF retinal ganglion cells cause different contrast sensitivities. *The Journal of Neuroscience*, 23(7), 2645-2654.

The Origin of Vierordt's Law

Zhuanghua Shi[1] and Stefan Glasauer[2]
[1]General and Experimental Psychology, University of Munich, Munich, Germany
[2]Computational Neuroscience, Brandenburg University of Technology, Cottbus-Senftenberg, Germany
shi@lmu.de; Stefan.glasauer@b-tu.de

***Abstract*—** In 1868 Vierordt discovered one type of errors in time perception – an overestimation of long durations and underestimation of short durations, known as Vierordt's law. Here we review the original study in its historical context. We found Vierordt's law is a result of an unnatural experimental randomization protocol, which Vierordt misused the "method of average error" that Fechner invented. Using iterative Bayesian updating, we simulated the original results with astonishing accuracy and concluded that Vierordt's law is caused by an unnatural yet widely used experimental protocol.

***Keywords*—** Vierordt's law, central tendency, Bayesian updating

I. INTRODUCTION

Karl Vierordt, professor of physiology at the University of Tübingen, in 1868 published his seminal book "Der Zeitsinn nach Versuchen" (Vierordt, 1868), which was the first quantitative attempt to investigate time perception with the methodology proposed and invented by researchers such as Ernst Weber, Gustav Theodor Fechner, and others. One of his main findings, and the one that best survived time, is now known as Vierordt's law. According to this law, short temporal durations tend to be overestimated, whereas long durations tend to be underestimated. Somewhere in between there is an "indifference point" where perceived time is veridical. The mechanisms underlying Vierordt's law have long remained obscure.

For his main experiments, most of them were done by Karl Vierordt as an only participant with the help of his assistant. Luckily, Vierordt explained his methods in detail and also published most of his data as tables. In the following, we concentrate on his Table A as an example (Vierordt, 1868, p. 36). It lists the average stimulus duration together with the signed error of reproduction for 22 intervals (from below 250 ms to above 8 s) and the corresponding number of repetitions (ranging from 25 to 83). The whole experiment consisted of overall 1104 trials presented consecutively, which clearly demonstrated the main feature of Vierordt's law, an overestimation of the short intervals and underestimation of the long intervals with the indifference point around 2.25 s (see Figure 1A).

The method used by Vierordt (Vierordt, 1868, p. 22) was the "method of average error" that Fechner invented (Fechner, 1860 Vol. 1, p120 ff, Vol. 2, p148 ff and p343 ff), now also known as method of adjustment. In Vol. 2 Fechner explained his method of average error in more detail (Fechner, 1860, p. 343). He applied 10 measurements of exactly the same condition (and same magnitude) consecutively. If there were multiple magnitudes, the magnitudes were tested in either increasing or decreasing order, and each magnitude was test in a chunk of 10 measurements.

A closer inspection of Vierordt's experiments shows various differences to the method proposed by Fechner. At least Vierordt partly knew that his method deviated from the one Fechner had proposed, but he defended those differences by claiming several advantages (e.g., p.29ff and p. 35 Vierordt, 1868). However, what is easily overlooked is that according to Vierordt in the experiments "the assistant provided … a time interval of arbitrary magnitude" (Vierordt, 1868, p. 35). According to Fechner's and Müller's descriptions, the method requires equal or ordered, rather than arbitrary magnitudes. Thus, evidently, the method used by Vierordt was not at all what Fechner had in mind.

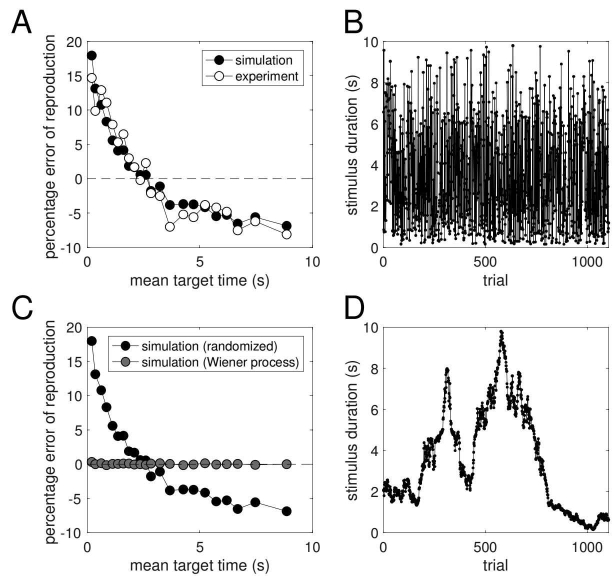

Fig. 1. Reproduction data of Vierordt's durations and iterative Bayesian models. (A) data from Vierordt's original experiment (open circles) and the best fitting model simulation (filled circles). (B) The sequence of the durations used for the simulation in A. (C) Comparison of simulation in A (filled circles) with the simulation from the sequence conforming to a random walk or Wiener process (gray filled circles). (D) The sequence of the durations for the simulation in C. Note that sampled durations in B and D are exactly the same, except for the sequential order.

Decades later, Woodrow aimed to replicate Vierordt's results but found no evidence for consistent over- and under-estimation in reproduced durations (Woodrow, 1930). Inspection of his methods shows that only one single interval was tested per day (50 repetitions). He explicitly mentioned: "Entirely different results might be expected from an experiment in which the various intervals were all employed on one day, particularly if they were used in an irregular order" (Woodrow, 1930). Thus, presenting the stimuli one by one and with sufficient temporal separation, as suggested by Fechner, apparently avoids the systematic errors that are the characteristic of Vierordt's law. In other words, Vierordt's law seems to be a consequence of the particular experimental protocol.

Over the next 80 years, various other investigations followed, but without providing a formal theory for Vierordt's law. In other fields of psychophysics, effects analogous to Vierordt's law were discovered for other types of magnitude estimation, such as "the law of central tendency" (Hollingworth, 1910), the "regression effect" (Stevens & Greenbaum, 1966), and the "range effect" (Teghtsoonian & Teghtsoonian, 1978). Interestingly, Hollingworth, who also referred to Vierordt's work, already provided important cornerstones of the effect, such as the indifference point depending on the range of stimuli given: "in all estimates of stimuli belonging to a given range or group we tend to form our judgments around the median value of the series" (Hollingworth, 1909). He concluded these remarkable insights from a series of experiments that he published in 1909, where he compared magnitude reproduction for different ranges of stimuli and for single stimuli presented in isolation (Hollingworth, 1909). Hollingworth's conclusions are thus providing evidence for the importance of the context of other stimuli in which a particular test stimulus is judged.

Following the notion of Hollingworth (1909), we hypothesized that if 1) Vierordt's law is a consequence of the experimental randomization, and 2) iterative Bayesian estimation can explain the central tendency, then we should be able to predict Vierordt's original data using the interactive Bayesian updating model (Petzschner & Glasauer, 2011, Shi, Church, & Meck, 2013) by applying the original experimental protocol as closely as possible. Moreover, the iterative Bayesian model also predicts Vierordt's law should be greatly reduced if the change of the magnitude is slowly and follows the random walk process.

II. Model simulation

Figure 1A depicts the Vierordt's original data together with the best fit from the simulation, and Figure 1B shows the best simulated sequence. Evidently, the model provides an excellent fit to Vierordt's data. However, how much does the reproduction error depend on experimental protocols? Supposing the same intervals are provided in ascending or descending order (assuming the same model with identical parameters), the model predicts the absolute percentage error would be below 0.2% for all intervals (as compared to below 15% in the original Vierordt's data). The differential outputs of the simulation corroborate our suspicion that Vierordt's law is a consequence of the random presentation of stimuli within the same experimental context. The iterative Bayesian updating model thus can explain both Hollingworth's conclusions about the central tendency and Woodrow's failure replication of Vierordt's findings.

Figure 1C illustrates this difference between a random walk sequence (Figure 1D) and a randomized sequence (note that stimuli in 1B and 1D are the same except for the temporal order of presentation). As we predicted, the central tendency was almost suppressed with the random walk sequence.

III. Conclusion

In summary, from a re-evaluation of the original dataset with iterative Bayesian modeling we conclude that Vierordt's law (and the central tendency) is a result of the specific experimental protocol - randomly presenting stimuli with large trial-to-trial magnitude fluctuation. This protocol deviates from what usually happens in everyday life, where either successive magnitudes are equal and share the same context, or different magnitudes are associated with different contexts. The proposed underlying mechanism of Bayesian dynamic updating indeed improves performance over trials for equal or slowly changing magnitudes but not for large magnitude fluctuations. According to our analysis, 150 years of research on Vierordt's law have thus focused on an effect that is caused by an unnatural but since then widely adopted experimental protocol, which was first introduced by Vierordt, who misinterpreted the method of reproduction invented by Fechner and described in his groundbreaking "Elemente der Psychophysik" (Fechner, 1860). After all, it was not Fechner's fault.

References

Fechner, G. T. (1860). *Elemente der Psychophysik*. New York: Holt, Rinehart and Winston.

Hollingworth, H. L. (1909). *The Inaccuracy of Movement: With special reference to constant errors*. The Science Press.

Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(17), 461–469.

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: a study on human path integration. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(47), 17220–17229.

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences*, 1–9.

Stevens, S. S., & Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. Attention, Perception & Psychophysics, 1(5), 439–446.

Teghtsoonian, R., & Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. Perception & Psychophysics, 24(4), 305–314.

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen*. H. Laupp'schen Buchhandlung.

Woodrow, H. (1930). The reproduction of temporal intervals. *Journal of Experimental Psychology*, *XIII*(6), 473–499.

Who Gives a Damn about Theory: Taten sagen mehr als Worte

Willy Wong
Department of Electrical and Computer Engineering and
Institute of Biomedical Engineering, University of Toronto

Abstract— Who cares about what a theoretical psychophysicist has to say. Empirical equations speak louder than theory! So here's an empirical equation for you. It's an equation concerning sensory adaptation. And it works universally.

Keywords— Adaptation, entropy theory, universality

I. Introduction

Psst… are you still reading? I thought we agreed that theoreticians have nothing interesting to say!

Well, ok. Perhaps you have decided that this the very last paper you will ever read from a theoretician.

Fact is, I'm a theoretician and even I find it difficult to penetrate other people's theoretical ideas! (Hahaha :-) But shhhh, let's keep this a secret, ok?)

So… why is it? Why do we ignore theoretical work, not just in psychophysics but also more broadly? Is it because it is difficult to penetrate? Is it because it's boring? Because it is not falsifiable?

I went to my first ISP meeting in 1994 in Vancouver, having the pleasure of meeting Lawrence Ward and many others for the first time. I was still a student then; it was my very first conference. Even back then, I started thinking about these questions.

Does theory make any impact whatsoever? And if it doesn't what could anyone possibly do to change it?

II. From Japan with Love I derive For You

I completed my graduate work with Ken Norwich working on the entropy theory. Some of you may remember Ken's contributions to ISP. After graduating and doing a postdoc in Japan with Shuji Mori, I eventually landed a faculty position at U of T teaching courses I never took as an undergrad. Talk about learning on the job. But I digress…

While at U of T, I continued to work on my doctoral research. In 2013, I got my first breakthrough 15 years after graduation and was happy enough to write a paper on this topic which was of course theoretical and mathematical in nature. I showed the paper to Shuji Mori, who by then was a very good friend and regular collaborator of mine. He was enthusiastic but tempered in his response about the paper. He said: "I think this is all good and fine, but really what is it telling us that is new in the world?" Well, he didn't really say that… I am just having a senior moment where I am losing both my short- and long-term memory.

But he had a point. It is easy to make theories and models consistent with what has already been found. It's a lot harder to predict something that hasn't yet been discovered or observed. Following William Whewell: "It is a test of true theories not only to account for but to predict phenomena."

III. The Psychophysicist with the Golden Equation

So, instead of telling you about my latest theoretical musings, instead, I offer you this. What if I told you there is an equation so simple that even high school students can comprehend it. An equation that works with sensory data as far back as the pioneering work of Edgar Adrian in his discovery of the all-or-nothing principle of action potentials. The equation governs experimental recordings spanning over a hundred years and is found to work in all of the sensory modalities (audition, vision, taste, touch, smell, head movements, etc.) and is obeyed in all kinds of animal species, including chordata (mammals, amphibians and fish), arthropods, molluscs and even cnidarians. It is a truly universal equation governing sensory function.

So, what is the equation exactly? Consider a usual adaptation curve. There is spontaneous activity (SR) prior to the stimulus onset, a peak response (PR) just after the stimulus is turned on, and eventually steady-state activity (SS) after some time:

$$\mathrm{SS} = (\mathrm{PR} \times \mathrm{SR})^{1/2}$$

That is, the steady-state is the geometric mean of the spontaneous and peak activities. This equation was predicted entirely theoretically from the entropy theory and exemplifies the highest example of simplicity and elegance in science.

Really? But I thought we already agreed that theoreticians have nothing interesting to say!

References

Norwich, K. H. (1993). *Information, sensation, and perception* (p. 247). San Diego: Academic Press.
Whewell, W. (1840). *The philosophy of the inductive sciences: founded upon their history* (Vol. 1). JW Parker.
Wong, W. (2020). *A Universal Law of Sensory Adaptation*. bioRxiv.