On Theories of Abstract, Quantitative Representation

by

Cory D. Bonn

Submitted in Partial Fulfillment of the

Requirements for the Degree

Doctor of Philosophy

Supervised by Professors Richard N. Aslin and Jessica F. Cantlon

Department of Brain and Cognitive Sciences
Arts, Sciences and Engineering
School of Arts and Sciences

University of Rochester

Rochester, New York

2016

Dedication

This dissertation is dedicated to my parents, John and Andrea Bonn, who have unwaveringly supported me through the many twists and turns of my academic career.

Biographical Sketch

The author was born in San Diego, California, USA. He attended the Eastman School of Music at the University of Rochester, Rochester, NY, and graduated with a Bachelor of Music degree in Applied Piano, a Take 5 Certificate in Linguistics, Psycholinguistics, and Philosophy in 2006, and a Master of Arts degree in Music Theory in 2010. Prior to the present doctoral study, he worked as a research assistant in the Department of Neurology at Massachusetts General Hospital in Boston, Massachusetts, in the laboratory of Dr. Mark J. Tramo and as a music instructor in the Boston area. He began doctoral studies in the Department of Brain and Cognitive Sciences at the University of Rochester in 2010 and subsequently received a Master of Arts degree in Brain and Cognitive Sciences in 2014. He was awarded a National Science Foundation Graduate Research Fellowship (NSF GRFP) in 2011, a NSF Graduate Research Opportunities Worldwide Fellowship in 2015, and a STEM-Chateaubriand Fellowship from the Embassy of France in Washington, D.C., in 2015. He pursued his research in infant methods, speech-category acquisition, and abstract quantity under the direction of Richard N. Aslin and Jessica F. Cantlon.

The following journal publications were a result of work conducted during doctoral study:

Bonn, C.D., & Cantlon, J.F. (2012). The origins and structure of quantitative concepts. *Cognitive Neuropsychology*, *29*(1-2), 149-173.

Stasenko, A., Bonn, C., Teghipco, A., Garcea, F.E., Sweet, C., Dombovy, M., McDonough, J., & Mahon, B.Z. (2015). A causal test of the motor theory of perception: a case of impaired speech production and spared speech perception. *Cognitive Neuropsychology*, *32*(2), 38-57.

The following are manuscripts with content from this dissertation to be submitted for publication:

Bonn, C.D., & Cantlon, J.F. (in prep). Spontaneous, modality-general abstraction of a ratio scale.

Bonn, C.D., & Cantlon, J.F. (in prep). Common cause or common effect? Comparing the relationships between pairs of quantitative dimensions.

Acknowledgments

I wish to thank my co-advisors, Dick and Jessica, for unyielding support through the twists and turns of my graduate-school career and providing the best possible models as mentors, scientists and collaborators. I also wish to thank Steve Piantadosi for the many impromptu conversations that challenged the core of my reasoning patterns and strengthened the way I reason about all kinds of problems, both scientific and non-scientific. I wish to thank Veronique Izard for taking me on as a new advisee and co-writing the grants that helped launch the race toward the dissertation finish line.

I would also like to thank the countless mentors I've had in other faculty members in the Department of Brain and Cognitive Sciences: Mike Tanenhaus, who served on my advisory and qualifying-exam committee and has, in no small part, shaped the way I think about cognitive science from the first undergraduate class I took in psycholinguistics to the grant-writing class I took in spring 2015; Robbie Jacobs, whose framing of issues in cognitive science, as well as direct comments, has influenced a great deal of the dissertation work; Celeste Kidd, who has shaped the way I think about developmental science and has been an ardent supporter since before I became a student in the department; Brad Mahon, who involved me in speech perception work in neurological patients and read through manuscripts; and Dave Knill, whose computational approach to science changed the way I frame research questions.

None of the research would have been possible without Kathy Corser, Jennifer Gillis, Chris Dambra, Chris Freemesser, Holly Palmeri, Rosa Li, Vy Vo, Courtney Lussier, and Katie Blakely, who always helped me get through issues small and large.

I would also like to thank faculty members outside the department who have been supportive of my career: Betsy Marvin, who has been an invaluable mentor and from whom I learned a great deal about teaching and music perception; Davy Temperley, who steered me in my first research directions as a music theory PhD student; Mark Tramo, a continued mentor, co-author, and collaborator in auditory perception and neuroscience, from whom I learned a great deal about rigorous scientific reasoning; Jeff Runner, who has been a constant supporter of mine since he advised my Take 5 program in 2005; Christine Gunlogson, who supported me in my Take 5 program, the beginnings of my interests in graduate school, and to the present day; and Nelita True, who, even as my piano teacher, encouraged my interests outside music performance, without which I would probably be on a totally different career path.

I would like to thank my friends and fellow graduate students in the department (present and past) who got me through the past five-and-a-half years with emotional and academic support: Katie Bankieris, Keturah Bixby, Sarah Bibyk, Esteban Buz, Dave Kleinschmidt, Jennifer Merickel, Ting Qian, Patty Reeder, David Ruskin, Sarah (Davis) Starling, and Ilker Yildirim. In addition, I would like to thank my close friends from the music-theory world: Sebastian Bisciglia, Sarah Marlowe, and Daphne Tan.

Finally, I would like to thank my family for their tremendous support: my parents, John and Andrea, my brothers, Geoff and Joe, and my sisters-in-law, Eva and Alison.

Abstract

Drawing on the possible connection between the similar psychophysical signatures of diverse, quantitative dimensions and a wealth of evidence showing that spatial, temporal, and numerical representations interact, Walsh (2003) proposed the existence of a central, generalized magnitude system in the brain. Though it has since been influential, stimulating a large body of research exploring its potential implications, its main claims are too imprecise to be falsified. In this dissertation, we develop a more precise characterization of the hypothesis space, drawing on literature from multisensory integration and learning as well as analogical reasoning. In a series of behavioral experiments in adults, we demonstrate that (1) caution should be exercised when drawing conclusions about a *generalized* system of magnitude representation from dual-response tasks because the evidence for a general computational solution for representing diverse pairs of magnitudes is inconclusive, and (2) a generalized magnitude representation may instead take the form of representations of relative magnitude that share a common dimensionless ratio, as observers can use ratio representations to compare sequences of magnitudes across sensory modalities and magnitude domains.

Contributors and Funding Sources

This work was supported by a dissertation committee consisting of Professors Richard N. Aslin (co-advisor), Jessica F. Cantlon (co-advisor), and Steven T. Piantadosi of the Department of Brain and Cognitive Sciences and Dr. Véronique Izard of the Laboratoire Psychologie de La Perception, Université Paris Descartes, Paris, France. The author collected and analyzed all data and wrote the manuscript for each of the chapters. Dr. Cantlon directly advised the experiments contained in the dissertation and served as co-author on published material from the introductory chapter. The majority of the text in the introductory chapter was published in a 2012 review article listed in the Biographical Sketch, reproduced here under the guidelines of publisher Taylor and Francis. Funding for the data collection was provided by NSF 1459625 and NIH R01HD064636 to Dr. Cantlon. The author was additionally supported by an NSF-Graduate Research Fellowship, NSF-Graduate Research Opportunities Worldwide Fellowship, and STEM-Chateaubriand Fellowship from the Embassy of France in Washington, D.C. Drs. Aslin and Piantadosi co-advised the present work and Dr. Izard advised the author on interpretations of results and ideas for future work.

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction.

Humans and other animals process information from many continuous psychophysical dimensions across sensory modalities and perceptual variables, including number, size, event duration, speed, visual and auditory brightness, pitch, and loudness. These dimensions constitute "quantities" or "magnitudes" because they are (at least roughly) interpretable as amounts, as in the case of number, size, and duration, visual brightness, and loudness, or because changes in stimulus values can be readily interpreted as amounts, as in the case of auditory brightness and pitch. Discrimination data collected from many such dimensions conform to Weber's law: successful discrimination of two stimuli along a given continuum depends on their ratio rather than absolute values. This is the principal signature of analog magnitude representations, in which values of increasing quantity are correlated with an increase in uncertainty (Gallistel, 1990; Gallistel & Gelman, 1992, 2000; Moyer & Landauer, 1967; Stevens, 1975; see also Cantlon, Platt, & Brannon, 2009, for review). This commonality in psychophysical performance suggests that, at some level, the neural computations required of each dimension are similar. In fact, a growing body of evidence suggests that different magnitudes—even those that might not be intuitively grouped at first glance—are related.

We begin by presenting the historical origins of currently discussed proposals for functional and neural architectures of magnitude representation. We then consider evidence that many dimensions of magnitude are related in the adult mind and brain, including space, time, number, pitch, and brightness. We include a brief overview and critique of the evidence discussed in previous reviews (Bueti & Walsh, 2009; Cantlon et al., 2009; Lourenco & Longo, 2011) as well as more recent, related findings; in addition, we include potentially related literature not previously discussed to illustrate the scope and complexity of the evidence to be explained.

We then briefly present non-mutually-exclusive theoretical frameworks within which it is possible to define abstract, quantitative representations and predictions of specific hypotheses. One is based on frameworks that have shaped investigations of multisensory integration and cross-modal transfer: the main idea is that causal structures in the world might provide correlated cues among magnitude dimensions. In the circumstance in which there are only a few different kinds of causal structures that generate those correlated magnitudes, a "generalized" computational solution, as proposed in previous literature, might be appropriate for the nervous system to implement. The other framework distinguishes the formation of abstract representations of quantity from the computational issues of cue integration and causality and only considers the problem of mapping from one dimension to another. These ideas serve as the motivation for the series of experiments described in the dissertation.

1.1 Origins of current debates

The question of how quantitative cognition is functionally organized in the brain began with neuropsychological investigations of patients with numerical and arithmetic impairments (e.g., Gerstmann, 1940). Neurological patients with parietal lesions can be impaired in making numerical judgments, while other cognitive abilities such as object categorization and recognition remain intact (Cipolotti, Butterworth, & Denes, 1991; Dehaene & Cohen, 1997). Furthermore, studies of patients with semantic dementia (and anterior temporal lobe atrophy) have shown that numerical skills can be spared in cases where other semantically demanding tasks such as picture categorization or picture naming are impaired (Cappelletti, Kopelman, Morton, & Butterworth, 2005). Those data show that it is possible to isolate numerical cognition from other components of cognition through damage to one (albeit large) part of the brain: the parietal cortex. Because those initial studies did not test magnitude representations for dimensions other than number, the question of whether the dissociation between "number" and other semantic domains is unique to numerical magnitudes cannot be resolved at present. In fact, it is not always the case that numerical deficits neatly segregate from other deficits: individuals with relatively focal lesions to intraparietal cortex commonly exhibit simultaneous deficits in arithmetic, spatial, and abstract perceptual judgments (Gerstmann, 1940; Takayama, Sugishita, Akiguchi, & Kimura, 1994). The functional relationships among those associated impairments have not been studied, and so it cannot be concluded that there is representational interdependence between arithmetic and visuospatial judgments. However, a long tradition of cognitive-science research supports the possibility that judgments of other perceptual intensities or magnitudes (e.g., size, time, brightness, loudness) could exhibit a similar pattern of impairment to numerical performance in these cases of neural impairment.

The idea that other magnitudes might share a common neural code was initially proposed by Gallistel and Gelman (2000). In their review of behavioral data from humans and other animals, they argued that discrete number should be represented with an analog magnitude code. Because animals must combine discrete number with continuous quantities in making decisions (e.g., in assessing food patches based on the number of potential food items and the space over which the food is spread), reconciling these incompatible formats necessitates conversion to a common code: the analog format.

Drawing on this suggestion, Walsh (2003) proposed that space, time, and other quantities—primarily number—share an abstract, undifferentiated magnitude code present at birth. His key claim is that an interconnected magnitude representation of time, space, and number emerges because of the critical role of magnitude information in the action system. The argument is that the common neural location of magnitude information and motor control in parietal cortex is what binds those computations. Although neural location could be an important factor in determining what cognitive representations are associated, a potentially more important factor is their functional origins in development. According to Walsh, the generalized magnitude system becomes differentiated in postnatal development, developing into specialized magnitude subsystems that share neural resources (in parietal cortex), though exactly how they are shared over development and to what extent each dimension is functionally differentiated remains unspecified.

One problem for understanding how magnitude dimensions are related is that a substantial amount of behavioral and neural evidence from humans and non-human animals is consistent with a number of theoretical possibilities for how magnitude relations developed or evolved to solve particular computational problems. These possibilities include innate relations, learned correlations, and both verbal/cultural and nonverbal analogies. We review the evidence for all three in the next section.

1.2 Inferences about the canonical domains of space, time, and number

Evidence for interactions among representations of space, time, and number comes from tasks that elicit representations in two of these dimensions simultaneously. In a now classic study of the interaction between time and number, Meck and Church (1983) found that rats are similarly sensitive to both number and duration (holding the other variable constant). In addition, they found that administration of methamphetamines increases the speed of the mechanism governing judgments of both dimensions, indicating that the animals' representations of time and number are subject to common constraints at some level of processing. Similarly, a recent study in human adults suggests that a click train can accelerate a common internal clock in sequential line, duration, and numerosity bisection tasks (Droit-Volet, 2010). Experiments in pigeons (Roberts, 1995; Roberts, Coughlin, & Roberts, 2000) and Stroop-like experiments in humans (Dormal, Seron, & Pesenti, 2006) have shown similar behavioral signatures (but see Roitman, Brannon, Andrews, & Platt, 2007, for evidence of a possible asymmetry in representations of number and duration). Taken together, these findings suggest that a common mechanism underlies judgments of both dimensions in these nonverbal tasks.

Reaction-time experiments also provide evidence of the interaction between number and space. In one early study, Henik and Tzelgov (1982) showed that when Arabic numerals are pitted against physical size, judgments of which numeral is larger (in size or number) showed congruency effects between the attended and unattended dimensions. Another classic signature is the so-called SNARC effect (Spatial–Numerical Association of Response Codes; Dehaene, Bossini, & Giraux, 1993). According to the most popular construal, adults map representations of number onto a horizontal mental number line, explaining the observation of faster processing times for larger numerical values on the right side of the line.

This type of mapping between space and number is also evident in cases of spatial neglect, wherein adult neurological patients with parietal cortex damage can exhibit asymmetries in their estimates of the "center" during both line bisection and numerical bisection tasks (Zorzi, Priftis, & Umiltà, 2002). Thus, the neural origin of space-number association appears to depend on parietal cortex. However, this is at least partly a learned association. The space-number mapping is known to emerge following exposure to counting behaviors and formal training in school (Berch, Foley, Hill, & Ryan, 1999; Opfer, Thompson, & Furlong, 2010; van Galen & Reitsma, 2008) and is flexible in bilinguals who learn two different spatial writing directions (Shaki & Fischer, 2008), strongly implicating culture in shaping this type of representation.

However, it is unclear whether the culturally mediated mapping of number onto space builds on a biological disposition to preferentially map number onto space (e.g., de Hevia & Spelke, 2010). One recent set of experiments suggests that space does not have a *privileged* psychological relationship with number in adulthood. Nuñez, Doan, and Nikoulina (2011) tested how well adult participants mapped number to non-spatial formats. In these experiments, participants mapped number onto non-spatial formats such as squeezing, bell striking, and vocalizing as well as spatial ones. Though the non-spatial mappings were found to be slightly different from the spatial mappings (they were logarithmically rather than linearly spaced and less precise), the authors suggest that there might not be anything biologically or conceptually special about the space–number relationship. However, an open issue is whether higher precision in the spatial judgments relative to non-spatial judgments is due to innate biases or extensive cultural experience with mapping dimensions onto space.

Much of the recent evidence for an interaction between representations of time and space in adults and animals comes from experiments that explore how language use might shape the development of abstract concepts. Across languages, the use of spatial language to describe time suggests that conceptualization of time is dependent on physical conceptions of space, though the exact way in which languages tend to conceptualize time in terms of space varies across cultures. For instance, according to Boroditsky (2000, 2001), Mandarin speakers are more likely to think about time in a vertical orientation than are English speakers, in line with the metaphors present in the language. Failures to replicate these results have rendered the original findings controversial (Chen, 2007; January & Kako, 2007; see Boroditsky, Fuhrman, & McCormick, 2011, for a response). Nonetheless, evidence from linguistic metaphors (e.g., Lakoff & Johnson, 1980, 1999) may suggest a link between space and time independently of whether the link is at a deep representational level or a superficial linguistic response level.

Results from tasks that presumably do not depend on language use suggest that the dependence of representations of time on space extends beyond the domain of language. Nonverbally presented magnitudes with no temporal component (i.e., static stimuli) modulate estimates of duration; larger, brighter, and more numerous stimuli are perceived to last longer than smaller stimuli of equal duration (Xuan, Zhang, He, & Chen, 2007; effects of Arabic numerals on duration estimation: Chang, Tzeng, Hung, & Wu, 2011). However, this effect is extremely small, and has come under heated debate in recent literature: some authors claim that reported interactions between space and time depend heavily on the type of response (Yates, Loetscher, & Nicholls, 2012; Rammsayer & Verner, 2014; 2015).

Eagleman (2008) suggests that these estimates of duration from size, brightness, and number directly reflect the amount of neural energy required for visual stimulus encoding, implying that the perception of interval duration is heavily influenced by aspects of stimulus encoding that are only indirectly related to timing (see also Pariyadath & Eagleman, 2007). The implications of those studies are not settled: either duration perception largely piggy-backs on the computational machinery of other magnitudes or the weight given to other magnitudes in overlapping representations is so large that it masks the input from true timing mechanisms. Under either interpretation, interactions between time, brightness, number, and size are fundamental.

A possibly related behavioral signature is an asymmetry of interference: in adult humans, judgments of line length interfere with judgments of duration more than duration judgments interfere with judgments of line length (Casasanto & Boroditsky, 2008). In one recent study, Merritt, Casasanto, and Brannon (2010) found that while adults' nonverbal judgments of duration are affected by the simultaneous representation of line length in a dual task, the effect of duration on judgments of line length are considerably smaller, consistent with previous findings. In rhesus macaques, duration and line length seem to interfere with each other equally, suggesting that the magnitude code the monkeys accessed is equally distributed between spatial and temporal representations. Thus there appears to be a spatially biased temporal representation in humans but not in monkeys. However, even though both the humans and the monkeys were trained to complete the task without verbal labels, human participants may still have linguistically encoded the durations (subvocally or otherwise) into English spatial terms (i.e., 'short' and 'long'). So, the uniquely human pattern of interference might arise at the level of lexical representation or response selection rather than a language-independent magnitude system. The results from the monkeys show that this asymmetry between time and size is a uniquely human phenomenon and therefore is unlikely to be purely a signature of interval timing as proposed by Eagleman (2008). Thus, we potentially can rule out the claim that the use of size, brightness, and number as a proxy for interval duration (described earlier) is the root cause of asymmetrical interference effects between space and time in humans.

The relationship between space and time has also been found at the neural level. Single-neuron data from neurophysiology studies of monkeys broadly indicate that the same neural substrate represents space, time, and number (Leon & Shadlen, 2003; Nieder, 2005; Roitman, Brannon, & Platt, 2007). Moreover, some data even suggest that a single parietal neuron can represent more than one type of magnitude. In one study, monkeys were trained to perform a line length-matching task and a numerical matching task (Tudusciuc & Nieder, 2007). During stimulus presentation or a subsequent delay, single neurons in ventral intraparietal cortex (VIP) responded selectively to visual stimuli based on their numerosity or length. Although some neurons responded only to numerosity and others only to line length, a subset of cells (20%) responded to the magnitudes of both the line lengths and the numerical values. In an adjacent parietal region, lateral intraparietal cortex (LIP), single neurons have been shown to be sensitive to quantities such as duration and number (Leon & Shadlen, 2003; Roitman et al., 2007). These and other studies, including fMRI studies of adults, have led some researchers to argue for a "distributed but overlapping" representation of different magnitudes at the neural level (Pinel, Piazza, Le Bihan, & Dehaene, 2004; Tudusciuc & Nieder, 2007). Moreover, Pinel et al. (2004)

found that the amount of functional overlap between brain regions recruited during Stroop-like tasks qualitatively predicted the size of the interference effects observed. Simply put, different types of magnitude representation including size, number, and time (and possibly brightness) share some neural resources in parietal cortex but not others.

Taken together, these findings tend to emphasize the relations between the dimensions of space, time, and number. The fact that there are so many studies that report a relationship between those dimensions (and not others) has led to arguments that there is a biologically privileged relationship among the dimensions of space, time, and number (Dehaene, Izard, Spelke, & Pica, 2008; Dehaene, Spelke, Pinel, Stanescu, & Tsivkin, 1999; Srinivasan & Carey, 2010; Walsh, 2003). However, as mentioned above, there is some evidence for fundamental interactions among quantitative dimensions beyond space, time, and number, such as interactions between time and brightness (Xuan et al., 2007). In the next section we review further evidence.

1.3 Beyond space, time, and number

There is some evidence that dimensions such as loudness, brightness, and pitch—dimensions other than those that are allegedly privileged (space, time, and number)—interact at the representational level. For example, adults are equally facile at scaling any kind of continuum to digital number as they are with scaling to loudness (Stevens, 1975). Furthermore, we already described evidence that among adult humans, brighter stimuli (in addition to larger and more numerous stimuli) are mistakenly estimated as lasting longer in duration than darker stimuli (Xuan et al., 2007). In addition, cross-dimensional mapping effects show up in Stroop-like paradigms for dimensions beyond space, time, and number. Marks (1987) showed that presentation of irrelevant, auditory-pitch information in visual brightness judgments (dark vs. light) and irrelevant, visual brightness information in auditory pitch judgments (low vs. high) affects adults' reaction times. In that experiment, irrelevant stimuli that were congruent (dark and low; light and high) facilitated responses, and those that were incongruent (dark and high; light and low) interfered with responses. In this section, we focus in particular on interactions found between non-canonical domains (e.g., auditory pitch and luminance) and canonical domains (number, space, and time).

1.4 Interactions with non-canonical domains

1.4.1 Luminance/brightness and loudness

Pinel et al. (2004) reported interactions between stimulus luminance and the canonical domains of number and space in magnitude comparison tasks. Irrelevant luminance information interfered with Arabic numeral and size comparisons, as indicated by a significant increase in response times on trials where the irrelevant dimension was incongruent with the relevant dimension. In addition, irrelevant information about physical size (but not number) interfered with luminance comparisons. The authors report that symmetrical interference only occurred between size and luminance. In combination with the finding that the amount of functional overlap in activation in parietal cortex parallels the size of interference effects, this study suggests that size, number, and luminance share computational resources and that, at least in adults, the representations of some pairs of magnitudes might be more closely related than others. Although it might not be intuitive that number and size should interact with luminance, one line of vision research suggests that they should: of two objects at equal depth, the brighter object will be perceived as closer (e.g., Farnè, 1977). This presumably is because the amount of retinal surface area stimulated by light reflected from an object increases as an object approaches. Thus, representations of subjective size and luminance may be linked to the perception of object distance.

Luminance interacts with perceived duration as well. With absolute duration held constant, humans, pigeons, and rats perceive bright lights as lasting longer than dim lights (Brigner, 1986; Kraemer, Brown, & Randall, 1995; Kraemer, Randall, & Brown, 1997; Wilkie, 1987; Xuan et al., 2007). Goldstone, Lhamon, & Sechzer (1978) report an effect of loudness on perceived tone duration as well as brightness on light duration in a magnitude comparison task. These results do not provide evidence for whether duration might also modulate brightness or loudness perception. However, at a broad level, these findings from pigeons, rats, and humans implicate fundamental interactions between time and brightness.

1.4.2 Pitch

A mapping similar to the SNARC effect occurs in the mapping of pitch height onto vertical space in adults (aptly dubbed the “SMARC effect”; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006). In this experiment, subjects were required to indicate whether a given pitch was lower or higher than a standard by pressing the space bar or the 6 key; each key stood for the “higher” response half the time. For example, pressing 6 for a “lower” response would result in longer response times, while pressing the same key for a “higher” response would result in faster responses. A similar signature of pitch–space representation has been observed in infants (Walker et al., 2010; though see Lewkowicz & Minar, 2014, for a replication failure in a slightly modified task and an alternative explanation based on perceived loudness). Mapping of pitch height onto vertical space in infants, musicians, and non-musicians alike may also result from peripheral filtering of the acoustic signal by the pinnae; shifting pitch height corresponds to shifting spectral peaks in the head-related transfer function (HRTF) for vertical space (Butler, 1969, 1971; Roffler & Butler, 1967; for an explanation relating these results to the HRTF, see Moore, 2003). Thus, the mapping of pitch height onto space might be explained at more than one level of representation—a consequence of learning the fundamental acoustic cues to object height or a higher-level, shared magnitude code.

Other evidence for the mapping of pitch onto a spatial representation, through either a common magnitude code or some other mechanism, comes from individuals with congenital amusia. Congenital amusia presents as impaired discrimination of fine-grained pitch changes and pitch-change directions (Ayotte, Peretz, & Hyde, 2002; Peretz & Hyde, 2003) and impaired short-term memory for nonverbal, auditory sequences (Tillmann, Schulze, &

Foxton, 2009; Williamson, McDonald, Deutsch, Griffiths, & Stewart, 2010). One study showed that amusic individuals' level of musical impairment on a standardized test correlated with impaired performance on spatial tasks (Douglas & Bilkey, 2007). In addition, impaired memory for changes in pitch-height direction reduces interference with spatial judgments in dual tasks in which both pitch and spatial judgments are made, suggesting that the representation of pitch height may share computational resources in individuals with typical pitch-processing capabilities (Douglas & Bilkey, 2007; but see Tillmann et al., 2010 for a replication failure in a lengthened version of the task).

Some studies also report interactions between duration and pitch. In the auditory kappa effect, tones are perceived as longer when their pitch is higher (Brigner, 1988; Cohen, Hansel, & Sylvester, 1954; see also Lourenco & Longo, 2011). In addition, the pitch difference between two tones increases both the perceived duration of the silence between them (Crowder & Neath, 1995; Shigeno, 1986) and the length of an intervening tone (Henry & McAuley, 2009). Moreover, amusics fail to show an auditory kappa effect at small pitch intervals (4 semitones; Pfeuty & Peretz, 2010). In the auditory tau effect, the pitch of the second tone in a three-tone sequence is affected by its timing; for example, when the middle tone is closer in time to the first tone, it is also perceived as closer in pitch (Christensen & Huang, 1979; Cohen et al., 1954; Henry, McAuley, & Zaleha, 2009; Shigeno, 1986). These studies indicate interactions between representations of pitch, space, and time.

Neuropsychological data that address the issue of neural overlap between representations of pitch and other magnitudes are currently sparse. Ideally, such tests would be administered within-subjects in cases of focal damage to a parietal locus that has caused a magnitude-related, performance impairment. One study that was conducted in that way showed that patients with damage to right posterior parietal cortex can exhibit impairments in judging the relative duration of two tones but remain unimpaired in judging a tone's pitch (Harrington, Haaland, & Knight, 1998). However, 2 out of the 10 patients in that study presented with comorbid impairments of pitch and duration judgment. The study shows that pitch and duration judgment are neurally dissociable processes in individuals with right posterior parietal lesions but that mutual impairment can occur in some of those cases. Interestingly, in patients with more anterior lesions, more than half of the individuals exhibited mutual impairment of pitch and duration judgments. Thus, pitch and duration judgments are commonly dissociated following posterior cortical lesions, but they are associated in cases of more anterior lesions. However, since the pitch judgment task was intended as a control task in that study, it is difficult to determine the nature of the association between duration and pitch judgments in the group with anterior lesions. Future studies would ideally include control tasks that allow performance to be functionally dissected.

1.4.3 Melodic contour

Ordinal relations among exemplars from one dimension are easily compared to ordinal relations from another dimension. In the domain of pitch height, ordinal relations among successive musical notes form melodic contour (Marvin & Laprade, 1987; Marvin, 1997). A more imprecise definition of contour is the "up" and "down" motion between successive notes (Dowling, 1978; Dowling & Fujitani, 1971). One study reports that similar contour relations may be found in other auditory patterns such as loudness, which is correlated with the amplitude of the waveform, and brightness, which is correlated with the portion of the frequency spectrum with the greatest concentration of energy (McDermott, Lehr, & Oxenham, 2008). Adults can match contours across auditory dimensions and recognize familiar melodies in dimensions other than pitch. The authors suggest that contour extraction is a generalized feature of auditory memory and may have a centralized processing mechanism. Prince, Schmuckler, & Thompson (2009) showed that adults can relate single visual line drawings to the contour of long melodies, suggesting that representation of contour may extend beyond the auditory modality. Further work is needed to assess whether contour extraction is based on ordinal computations from a generalized analog magnitude system such as the one that may underlie the processing of ordinal relations for space, time, and number. The representation of ordinal relations, including relative judgments of number, size, time, loudness, brightness, and pitch (i.e., mental comparisons) might be functionally interdependent and share mechanisms, or they could be functionally parallel and rely on mechanisms that are not shared but, rather, operate in a similar way. In fact, this conclusion could also apply to much of the data that show associations among dimensions such as number, space, and time: the data are often ambiguous as to whether their relationship is one of functional interdependence or a functional parallel.

1.5 Summary of evidence from human adults and animals

Based on the above evidence, several open but empirically tractable issues can be identified. One issue is the extent to which the representations of different magnitudes and their associated computations are distinct. Another issue is the rather unconstrained space of what more comprehensive descriptions of the functional architecture of the magnitude system(s) should and could look like, given the unspecified set of computational problem(s) the world poses. A final issue is at what level of representation the observed interference effects arise and whether asymmetries of interference reflect unequal distribution of computational resources, something deeper about the architecture of the magnitude system, or idiosyncratic internal models specific to each pair of magnitudes. Most extant studies are consistent with many views of magnitude relations, including innate constraints on certain magnitude relations or any number of statistical learning models.

1.6 Developing theoretical frameworks to explain the existing data

In order to understand what taxonomic distinctions exist among magnitudes, it is important to consider how associations among magnitudes might originate in the mind and brain and therefore what their functional relations could be. In the following section, we consider existing ideas about magnitude representation within the larger context of theories of abstract percepts and concepts in development and adulthood. We extend frameworks from cross-modal and multisensory perception to the conceptual structure of abstract relations among magnitudes and contrast it with an alternative framework distantly related to theories of analogical reasoning. The frameworks presented here are not mutually exclusive and could even be complementary.

1.7 Frameworks for understanding relationships between magnitude domains

Little is known about the development of generalized magnitude representations. We first borrow hypotheses and evidence from research on the development of multisensory perception to help frame research questions concerning the relations between magnitudes in the natural environment. This approach is justifiable because multisensory representations of abstract percepts (e.g., object location, typically modeled as a weighted combination of visual and auditory cues) are similar to generalized magnitudes in the sense that information from separate sources can be bound together and/or influence composite representations that factor in information from more than one source.

1.7.1 Types of relationships among magnitudes

One intuitive conceptualization of a generalized magnitude system is that it arises as a composite representation of correlated magnitude data provided by the environment (eg., de Hevia, Izard, Coubart, Spelke, & Streri, 2014). The implicit assumption is that causally related sources of sensory stimulation provide correlated cues in the environment, and thus the correlated sensory cues could be used to make inferences about latent (or distal) environmental causes. Inferences and predictions could be made (1) by constructing something like a regression model, in which connections between pairs or groups of magnitude arise as a simple, monotonically increasing function, or (2) by reference to more flexible, higher-level, representations of abstract features. In the multisensory-integration and cross-modal perception literature, these higher-level representations are often referred to as amodal because they refer to a property that does not belong to any one sensory modality alone. Both types of representation may be present at birth or built and tuned across development. We describe them in greater detail below.

1.7.2 Building internal models via associative learning

The binding of correlated events across senses and cognitive domains (assumed to be innately separate) has been the historically dominant explanation of the development of multisensory percepts and abstract concepts (e.g., Piaget, 1952, 1954). At a minimum, an associative account requires that infants and children construct representations of correlations among percepts via a kind of direct connection: stimulation of one dimension at a certain magnitude leads to activation of a specific distribution over values of magnitude in another dimension (a likelihood function), and this mapping is refined on the basis of previous observations from some estimation procedure that minimizes error. As we alluded to earlier, postnatal learning need not be the only process that explains the construction of such mappings; infants may come with innate, prior expectations about any given likelihood function. Asymmetries in the influence of one magnitude on another fall out of these types of models easily: the likelihood $p(X \mid Y)$ does not equal $p(Y \mid X)$.

1.7.3 Latent-variable magnitude representations

A more sophisticated strategy than a simple, associative-learning model or univariate regression model would be one in which the observer maps multiple, redundant sensory measurements onto an additional, unobserved quantity. Such a variable could be like a latent cause; these kinds of models have been successful in explaining optimal integration and segregation behavior in multisensory processing (Körding, Beierholm, Ma, Quartz, Tenenbaum & Shams, 2007); in addition, the construction of sets of latent features via non-parametric, Bayesian models has been successful for explaining the acquisition of latent multisensory features (Yildirim & Jacobs, 2012) as well as abstract feature sets for explaining similarity judgments among objects (Austerweil & Griffiths, 2011; 2013).

The multisensory integration literature has a deep connection to a much older body of evidence that suggests that amodal representations of multisensory inputs (including magnitudes) exist early in development. Eleanor Gibson (1969) sought to explain cross-modal matching behavior in multiple domains (e.g., for intensity across several modalities, Stevens, Mack, & Stevens, 1960; for higher-order figural properties, Rudel & Teuber, 1964; and for case studies of letter identification after vision restoration, Gregory & Wallace, 1963) and cross-modal transfer (e.g., for non-conventional shape, Caviness, 1964). In this view, an abstract, amodal representation of intensity or amount of stimulation is present from birth or very early in infancy and thus represents an innate component of multisensory perception. Gibson thought of amodal representations as dividing into two possible types, both of which rely on information redundancy. Her discussion of amodal relations includes two types: (1) inter- sensory redundancy (e.g., timing information about hammer strikes can be sampled from both the auditory and visual modalities) and also (2) relative intensity (e.g., "sharpness, bluntness, and jerkiness"; Gibson, 1969, p. 219).

Since Gibson, evidence for amodal representations in infancy has come from demonstrations of information transfer across modalities in infants (from oral to visual, Gottfried, Rose, & Bridger, 1977; from tactile to visual, Gottfried et al., 1977; Meltzoff & Borton, 1979; though see Maurer, Stager, & Mondloch, 1999) as well as demonstrations of cross-modal equivalence (e.g., continuity/discontinuity and ascendancy/descendency, Wagner, Winner, Cicchetti, & Gardner, 1981). Across those demonstrations of amodal representation, the transfer of information from one modality to the next is not necessarily equally strong in either direction (e.g., in visual–tactile transfer, Bushnell & Weinberger, 1987).

1.7.4. Abstract representations of relative magnitude: a more general solution?

The previous frameworks possess an important limitation in that the types of representation are only as general as the computational problems they solve. It may be the case that correlations among magnitudes (or any kind of relationship between magnitudes) arise from heterogeneous types of causal structures, precluding any rational solution that would assume any type of generality. It may also be the case that the nervous system irrationally or inappropriately recycles the same computational solution across evolution or development for a heterogeneous set of problems (for a description of a rational rather than irrational kind of reuse of preexisting neural machinery, see Dehaene & Cohen, 2007).

In addition, the previous characterization of amodal representations of magnitudes across the senses is ambiguous: are these representations of *absolute* magnitude levels or *relative* magnitude levels? If they are relative, then there is the possibility that a generalized magnitude system could arise naturally from the calculations that give rise to the relative magnitude representations rather than the need to solve a causal-inference problem or calculate a latent quantity by combining multiple inputs. Ratio representations and ordinal representations of values along a continuum are naturally dimensionless—meaning they are not grounded by any particular reference metric. Moreover, subjects spontaneously represent proportions within dimensions such as length, numerosity and fractions in Arabic numerals (Vallentin & Nieder, 2008; 2010; Jacob & Nieder, 2009; Jacob, Vallentin, & Nieder, 2012). Thus, this could be a type of 'common code' shared by all magnitude domains that produce relational information. Mappings between these representations would represent a sophisticated form of analogical reasoning rather than a solution to the problem of *binding* different magnitudes together with a likelihood function or a latent-cause representation. Structures could be analogically similar without being bound together in a representation of a distal, environmental cause, but the two are not mutually exclusive (see Srinivasan & Carey, 2010, for a related discussion of *structural similarity*).

1.8 Overview of the dissertation

In this dissertation, we aim to evaluate the following, given the presented frameworks: (1) the extent to which behaviors related to the simultaneous representation of magnitude values recruit generalized or specific, computational solutions, and (2) the possibility that a candidate for a generalized magnitude code could be a (previously underemphasized) representation of relative magnitudes. The first two chapters assess magnitudes via naturally arising behaviors with no training. The second two involve supervised statistical learning to further probe how adults may construct internal models of abstract magnitudes.

Chapter 2 presents exploratory work in which we systematically test all possible pairs of a subset of magnitude dimensions in a dual-production task and ask whether any patterns arise in the inter-dimensional biases that would suggest the use of a generalized computational solution. The results fail to reveal any patterns of inter-dimensional influence that would suggest a generalized computational solution.

Chapter 3 demonstrates that subjects can use representations of ratios (and ordinal relations) among items in pairs of sequences to rate their similarity. These comparisons are demonstrated within sensory modality and across sensory modality (and magnitude domain), suggesting a truly abstract, spontaneous representation of ratios that can be used to map magnitudes from one domain to another.

Chapter 4 uses a supervised statistical learning paradigm to show that adults can transfer fine-grained representations of stimulus distributions across sensory modalities, though the distinction between absolute and relative magnitudes is not made here. The result could reflect both a sophisticated type of analogical reasoning or the expectation that magnitude values are correlated across causally related sensory inputs.

Chapter 5 uses a different, supervised statistical learning paradigm to probe the structure of the internal model that adults create when forced to learn that two magnitudes are correlated. In addition, we show that two types of inter-dimensional relationships commonly cited for a generalized magnitude system, (1) the relationship between size and duration as exhibited by the previously discussed bias effects, and (2) the SNARC effect, which demonstrates a relationship between numbers and left-right location (space), may actually draw on different kinds of internal models. In other words, subjects bring different models to the same statistical learning task when confronted with two different pairs of magnitude dimensions or continua. At the very least, this confirms the need for caution when drawing conclusions about the existence of a *generalized* magnitude system arising from the need to explain correlated magnitude values from pairs of dimensions in the environment.

In the concluding chapter, we recapitulate the findings from the presented series of experiments and probe further possibilities for why different kinds of magnitudes may interact, focusing specifically on space, time and number.

Chapter 2. Bias effects in dual-magnitude production tasks: Pairwise comparisons of size, duration, number, and brightness

2.1 Preview

The extant literature on generalized magnitudes, particularly literature that uses dual-attention tasks to elicit inter-dimensional bias effects, suggests that—at the very least, a systematic pattern of biases should emerge in systematic, pairwise testing. In a series of 6 experiments using a dual production task, we probed this possibility in pairwise testing of the dimensions of size, numerosity, duration, and brightness. No clear patterns of influence emerged, though several unidirectional effects were revealed—some unexpected, given extant literature. The results suggest that, at least for magnitude production tasks, performance in any given pair of domains reflects idiosyncratic interactions rather than systematic effects of a generalized magnitude system. However, this does not preclude the possibility that a generalized system of a response code may arise in tasks requiring categorical or ordinal responses.

2.2 Introduction

The largest source of evidence for a generalized magnitude system in adults comes from dual-task studies in which subjects must attend to two magnitude dimensions within or across sensory modalities simultaneously for a given stimulus (see Bueti & Walsh, 2009; Bonn & Cantlon, 2012, for reviews). Less frequently, authors will require only monitoring of one domain and report much smaller effect sizes (eg., Chang, Tzeng, Hung, & Wu, (2011); Rammsayer & Verner, 2014; 2015). Typically, stimulus magnitudes in one domain will interfere with or bias another in a manner similar to the Stroop task (1935).

However, it is unclear whether bias or interference effects are the result of a generalized magnitude system: it is difficult to extract overarching patterns across the current published literature, as in a meta-analysis, with the variety of task manipulations and stimulus levels tested. Moreover, there is likely a bias for reporting positive findings (file-drawer effect resulting in a biased pool of literature) from proponents of the inter-dimensional bias hypothesis.

Independently of publication bias, it may be inappropriate to draw conclusions about an abstract magnitude system from Stroop-like tasks in general. It could be the case that inter-dimensional bias effects, by virtue of being examined for presence or absence, reflect a processing mechanism that is idiosyncratic to each pair.

To examine this possibility, we sought to systematically test inter-dimensional bias effects in a series of pairwise comparisons of four of the key, visual magnitude domains cited in the literature: object size, numerosity (dot arrays rather than Arabic numerals), duration, and brightness. If a generalized magnitude system exists and is responsible for inter-dimensional bias effects, then a systematic pattern of inter-dimensional bias should emerge. For example, one domain like size may asymmetrically influence the others (as in Casasanto & Boroditsky, 2008; Merritt et al. 2010).

2.3 Overview of Experiments

A wealth of literature using diverse methods has documented the biasing effects of different magnitude domains on each other. For instance, under certain circumstances, size perception is known to influence duration judgments (Casasanto & Boroditsky, 2008; Merritt, Casasanto, & Brannon, 2010; Xuan, Zhang, He, & Chen, 2007; Rammsayer & Verner, 2014; 2015).

However, a great deal of variation exists across tasks, the range of magnitude values used within experiments, as well as the number of times different pairs of magnitudes have been tested together. In the following set of 6 experiments, we asked subjects to perform the same, dual-production task on the complete set of pairs of magnitudes from the visual domain: object size, stimulus duration, stimulus brightness, and numerosity (dot arrays). The experiments were primarily exploratory: the purpose was to compare inter-dimensional bias effects across these pairs of magnitude on a level playing field. Secondly, if a generalized magnitude system governs interactions between the dimensions, then systematic patterns should arise in the data; for example, pairs of dimensions may cluster together to mutually influence each other.

2.4 General Method

2.4.1 Subjects

Subjects were recruited on Amazon Mechanical Turk (www.mturk.com) and restricted to having a 95% MTurk approval rating and IP address within the

United States. We recruited 16 subjects for each experiment and they were compensated $6 for 45 minutes of work.

2.4.2 Stimuli

Stimuli were created in the web browser using HTML Canvas and the fabric.js library (www.fabricjs.com; Zaytsev & Chernyak, 2014). The viewing area was an 800-by-600 black canvas with a 1-pixel, white border.

Subjects controlled experiment flow and certain kinds of production responses with button icons on the left-hand side of the screen. In addition, to help subjects keep track of the flow of the experiment, trial tallies were displayed just above the button control panel.

Stimulus values were drawn from an 11 × 11 matrix of values: each of 11 logarithmically spaced magnitude values from one dimension was paired once with each of 11 values from another.

2.4.3 Procedure

After pressing the ‘Begin Trial’ button, subjects observed a single visual stimulus within the viewing window at a random location determined online. A 1000-ms delay occurred before presentation of the production prompt in the center of the viewing area.

At that time, subjects had to reproduce the magnitudes on each of the 2 tested dimensions that they observed, using custom responses tailored to each magnitude dimension. No training or catch trials were included.

2.4.4 Exclusion of Subjects

Subjects' whose reproduced values on each of the two tested responses were not significantly correlated with the presented stimulus values were excluded from the analysis, as this indicated lack of attention or effort.

2.5 Method: Experiment 1, Size-Duration

2.5.1 Stimuli

Stimuli were filled circles (light-blue hexadecimal color #33CCFF to prevent fatigue from looking at bright dots on a black background) presented with the following radii (in pixels) and durations (in ms). One pair was presented on each trial from the following list, with all possible combinations presented twice. *Radii:* {30, 34, 40, 45, 52, 60, 69, 79, 91, 104, 120}; *durations:* {750, 862, 990, 1137, 1306, 1500, 1723, 1979, 2274, 2611, 3000}.

2.5.2 Procedure

Subjects observed the stimulus, and then were prompted to reproduce both the size and duration of the stimulus by manipulating a generic dot sized at the geometric mean of the set of radii (60-pixel radius). Subjects could record their desired size or duration in any order they wished.

To record their size response, subjects could drag one of four corner handles provided by the fabric.js library to adjust the size of the circle. To record a duration production, subjects pressed a button labeled 'Record Duration.' After this button was pressed, the text on the button changed to read "RECORDING."

Subjects were instructed to click and hold the dot for as long as the remembered duration of the stimulus. As they clicked and held the dot, the dot changed color to red to show subjects that the recording process was working. When they released the mouse button (or trackpad), the dot turned back to #33CCFF blue.

Because unexpected browser events can disrupt timing recording, we also provided a means for subjects to preview their response in all experiments involving duration memory in case subjects suspected their estimated duration had been inaccurately recorded. They could click 'Preview My Response' to see their created stimulus.

To indicate that they were finished recording, subjects clicked on a 'Finished' button to commit their response.

2.6 Method: Experiment 2, Size-Numerosity

2.6.1 Stimuli

Stimuli were arrays of blue dots (hexadecimal #33CCFF) that varied in numerosity enclosed in a blue circle that varied in size. The average size of the blue dots inside the circle increased in proportion to the area enclosed by the blue circle, precluding most density cues to numerosity caused by the varying size of the overall display. At most, the spatial extent of the dots was 50% of the area enclosed by the circle on any given trial. The brightness values of all the dots in the array were manipulated by randomly choosing an opacity value from a uniform distribution between 50% and 100%. Lastly, the dots were constrained to be non-overlapping.

Displays could be of the following circle sizes and numerosities. *Sizes:* {50, 57, 66, 76, 87, 100, 115, 132, 152, 174, 200}; *numbers:* {7, 8, 9, 11, 12, 14, 16, 18, 21, 24, 28}.

2.6.2 Procedure

Following the post-stimulus interval of 1000-ms, subjects were presented with a display at the geometric means: a circle of a 100-pixel radius enclosing a dot array with 14 dots. Subjects could increase or decrease the number of dots by pressing the 'Add Dot' or 'Subtract Dot' buttons on the control panel. They could increase or decrease the size of the outer circle with one of four corner handles provided by the fabric.js library. Crucially, the size and spacing of the dots in the numerosity display increased or decreased in real time in proportion to the size of the outer circle as subjects adjusted it. When subjects were finished recording their productions, they clicked 'Finished' to end the trial.

2.7 Method: Experiment 3, Size-Brightness

2.7.1 Stimuli

Stimuli were filled circles of the same sizes as in Experiment 1, Size-Duration. To maximize the range of perceived brightness, these dots were white rather than blue, contrasting with the previous two experiments. To manipulate perceived brightness, we adjusted the opacity level of the dots. This ensured more reliably constant judgments of brightness across different monitors, as different monitors have varying levels of inherent brightness and contrast values that cannot be controlled. The dots could be of the following opacity values: {10%, 13%, 16%, 20%, 25%, 32%, 40%, 50%, 63%, 79%, 100%}.

2.7.2 Procedure

Following the post-stimulus delay, a new dot at the geometric mean size 60-pixels and 50% opacity appeared. Subjects could adjust the size of the dot by clicking one of four corner handles, again provided by fabric.js. To adjust the opacity of the dot, subjects used a horizontal slider below the viewing area that was 400-pixels wide, with the leftmost value equaling 10% opacity and the rightmost value equaling 100% opacity. The slider handle always started at the middle. For simplicity of calculation, the values on the slider were linearly scaled. The words 'Dimmer' and 'Brighter' appeared on the left and right sides of the slider, respectively. The brightness of the dot was adjusted in real time in response to observer input.

As with the previous experiments, subjects could manipulate their responses at their leisure until clicking 'Finished.'

2.8 Method: Experiment 4, Duration-Numerosity

2.8.1 Stimuli

Stimuli were dot arrays that followed the same constraints as in Experiment 2, except the outer circles were not displayed to the subjects. These displays were on screen for the same durations presented in Experiment 1.

2.8.2 Procedure

After stimulus presentation, a new display at the geometric mean number of dots appeared inside an inconspicuous blue square. Numerosity productions were recorded using the same method as Experiment 2 (with 'Add' and 'Subtract' buttons). Duration productions were recorded in the same manner as Experiment 1, though subjects could click anywhere in the vicinity of the dot display to record their duration after pressing the 'Record Duration' button. Instead of all the dots turning red during recording, the blue border turned red for the duration of a click. This was to save computing resources during rendering.

.

2.9 Method: Experiment 5, Duration-Brightness

2.9.1 Stimuli

Stimuli were 60-pixel-radius, white dots that appeared for one of the durations used in Experiment 1 and 4 and at one of the opacity levels used in Experiment 3.

2.9.2 Procedure

Subjects manipulated duration in the same manner as in Experiment 1 and manipulated opacity as in Experiment 3.

2.10 Method: Experiment 6: Numerosity-Brightness

2.10.1 Stimuli

Stimuli were arrays of white dots that appeared for 1500 ms at one of the opacity values used in previous Experiments 3 and 5 and numerosity values from Experiments 2 and 4.

2.10.2 Procedure

Subjects manipulated numerosity as in previous experiments with 'Add' and 'Subtract' buttons and manipulated brightness with a slider below the viewing area.

2.11 Results

We first summarize the results of multilevel regression models (Gelman & Hill, 2007) designed to test for the effects of one stimulus value on the other in reproduction. Our general strategy was as follows: we fitted a linear regression model with a full set of random effects (slopes and intercepts by subject) and estimated significance of effects using the Kenward-Roger approximation for degrees of freedom (Kenward & Roger, 1997). In the event that model estimation failed due to convergence issues, the random slope for the main dimension was removed to retain maximally conservative estimates for the secondary dimension. Specific occurrences are indicated with asterisks in the regression table, Table 2.1.

For the fixed effects, we entered two: the stimulus value on the same dimension as the response value (eg., stimulus duration predicting response duration) and the stimulus value on the simultaneously presented dimension. Thus, the effect of the simultaneously presented dimension was assessed while controlling for the effect of the main dimension, assessing any cross-dimension bias effects. The results of the tests are shown in Table 2.1.

There were a few results suggesting small, but significant levels of inter-dimensional influence. There were only three significant results among the 12 possible inter-dimensional effects: duration predicted size judgments ($p = 0.003$), size influenced numerosity judgments ($p < 0.00001$), and numerosity influenced brightness judgments ($p = 0.0117$). There were two marginal effects: an effect of numerosity on size ($p = 0.0692$) and an effect of brightness on size ($p = 0.0768$). Of all these effects, the strongest was the effect of size on numerosity judgments ($B = -5.93$); moreover, this was the only *negative* biasing effect: the larger the size, the smaller the numerosity estimate.

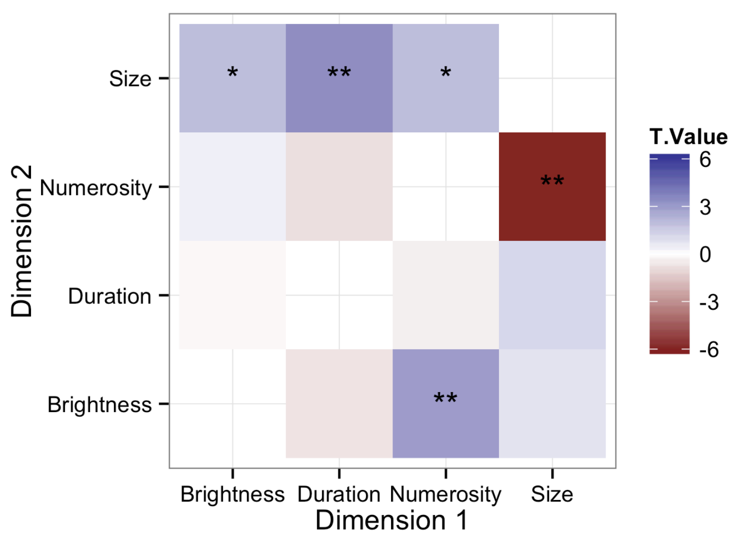
No single experiment yielded any reciprocal influence effects; in every case of significant or marginal effects, only one dimension influenced the other. Two of the 6 experiments failed to yield any inter-dimensional influence at all: the duration-brightness experiment and the duration-numerosity experiment. A summary of the effects is shown in Figure 2.1.

Table 2.1. Results from regression analyses for all 6 dual-production experiments. Significance testing was performed using the Kenward-Roger approximation of the degrees of freedom. Significant results reflecting bias from the non-response dimension to the response dimension highlighted in dark gray. Marginal results highlighted in light gray. Significant results reflecting the effect of the reproduced stimulus dimension are not highlighted but indicate subjects successfully reproduced values in the requested domain. Asterisks indicate one removed random slope for the main stimulus dimension for facilitating model convergence.

EXPERIMENT	*B*	*SE*	*t*	*p*
Experiment 1, *n* = 15				
*Size Predicting Duration**				
(Intercept)	506.56	69.33	7.31	0.0000
Duration	0.66	0.01	45.06	0.0000
Size	0.42	0.37	1.14	0.2691
*Duration Predicting Size**				
(Intercept)	5.72	1.48	3.86	0.0010
Duration	**0.00**	**0.00**	**3.33**	**0.0034**
Size	0.83	0.01	95.04	0.0000
Experiment 2, *n* = 14				
Size Predicting Numerosity				
(Intercept)	4.99	1.05	4.75	0.0004
Size	**-0.01**	**0.00**	**-5.93**	**0.0000**
Numerosity	0.75	0.07	11.43	0.0000
Numerosity Predicting Size				
(Intercept)	24.06	3.54	6.80	0.0000
Size	0.71	0.04	17.71	0.0000
Numerosity	**0.12**	**0.06**	**1.98**	**0.0692**
Experiment 3, n = 15				
Size Predicting Brightness				
(Intercept)	0.13	0.01	17.30	0.0000
Size	0.00	0.00	0.78	0.4355
Brightness	0.71	0.01	92.29	0.0000
Brightness Predicting Size				
(Intercept)	4.27	1.66	2.57	0.0223
Size	0.90	0.03	27.43	0.0000
Brightness	**1.45**	**0.76**	**1.91**	**0.0768**
Experiment 4, n = 13				
Duration Predicting Numerosity				
(Intercept)	2.18	0.35	6.30	0.0000
Duration	0.00	0.00	-0.86	0.4014
Numerosity	0.83	0.03	28.19	0.0000
Numerosity Predicting Duration				
(Intercept)	527.12	66.65	7.91	0.0000

Duration	0.61	0.07	8.39	0.0000
Numerosity	-1.13	2.63	-0.43	0.6733
Experiment 5, n = 14				
Brightness Predicting Numerosity				
(Intercept)	2.97	0.79	3.74	0.0025
Numerosity	0.76	0.05	14.15	0.0000
Brightness	0.42	1.04	0.41	0.6921
Numerosity Predicting Brightness				
(Intercept)	0.17	0.03	5.12	0.0002
Numerosity	**0.00**	**0.00**	**2.93**	**0.0117**
Brightness	0.65	0.06	11.64	0.0000
Experiment 6, n = 13				
*Duration Predicting Brightness**				
(Intercept)	0.13	0.01	9.62	0.0000
Duration	0.00	0.00	-0.72	0.4732
Brightness	0.77	0.01	90.35	0.0000
Brightness Predicting Duration				
(Intercept)	443.24	133.26	3.33	0.0060
Duration	0.80	0.06	12.64	0.0000
Brightness	-17.81	113.32	-0.16	0.8777

Figure 2.1. Summary of cross-dimension effects (in *t* values) across the 6 experiments. X-axis represents the influencing dimension and the y-axis r represents the influenced dimension. Colors represent *t* values so that all effects are on the same scale, with blue shades being positively signed and red shades being negatively signed. White cells (representing within-dimension influence) on the diagonal do not contain values. Stars represent significance level: ** = significant and * = marginal.



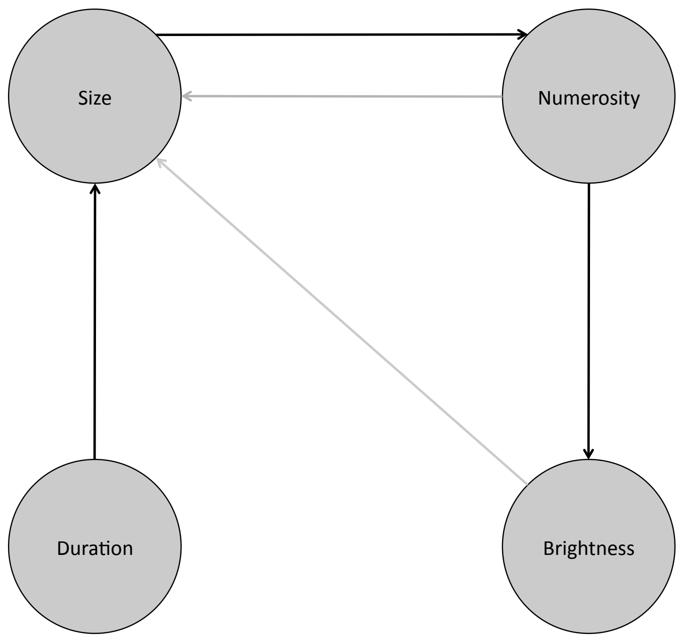
2.12 Discussion

The effects reveal a patchwork of surprisingly diverse results. Numerosity, duration, and brightness all influenced size (if one includes marginal results for numerosity and brightness), but size only influenced numerosity judgments *in the opposite direction*. In addition, numerosity influenced brightness judgments. Importantly, no symmetrical effects were found. Incongruent with previous literature is that there was no influence of size on duration judgments, even though there was an influence of duration on size judgments. This indicates an asymmetry that is the *opposite* of that found in the literature in which duration influences size more than size influences duration (eg. Casasanto & Boroditsky, 2008). In addition, the study failed to reveal an effect of numerosity on duration judgments, in contrast with Chang et al. (2011), who used Arabic numeral stimuli in a duration-only response task. Figure 2.2 displays a graphical summary of the directed effects.

What do these effects reveal? While the experiments may have yielded results that are idiosyncratic to the particular method or stimulus set, that is also a possibility with all previous results in the literature. In any case, it is important to note the failure of the generality of effects across these experiments and that while inter-dimensional influences do exist, they are far from ubiquitous or even patterned in a way that would suggest the existence of a generalized magnitude.

Instead, the results suggest that each pair of dimensions likely interacts in a unique way and that each interaction merits a unique space of hypotheses concerning its potential causes. This, however, does not preclude the existence of a generalized magnitude system: it simply suggests that literature using inter-dimensional biasing effects may not be the most appropriate evidence for one.

Figure 2.2. Graphical summary of dual-task bias effects. Arrows indicate direction of influence. Gray lines represent marginal effects and black lines represent significant effects.



In the following section, we consider explanations for each effect individually.

2.12.1 Effects of duration

Why would duration affect size judgments but not brightness or numerosity judgments? It may be that in the natural world, size and duration share a causal relationship but duration, numerosity and brightness do not. However, it remains unclear what computational problem the mind is solving that yields a relationship between duration and size if there is no effect of size on duration judgments.

2.12.2 Effects of size

Why does increased size decrease numerosity judgments? The stimuli equated the proportional size of the each dot for every possible size of the enclosing circle and, in probe stimuli, the inter-dot distance increased proportionally as subjects manipulated display size, so the proportional cumulative surface area for each number for each size was equated, obviating overall density as the explanatory factor. However, it could be that subjects have a prior bias to expect fewer objects in a scene when the individual objects are larger, independently of inter-item spacing.

2.12.3 Effects of numerosity

Why does increased numerosity increase reproduced brightness (opacity)? It could be that subjects expect more objects to produce more total light in these scenes. However, *subjective* brightness was not controlled for in the total number of objects, so it could be that individual scenes with more objects were subjectively brighter than scenes with fewer objects and this effect was exaggerated in working memory.

Why would increased numerosity marginally increase estimated size? One intuitive possibility is that subjects may bring a prior expectation of greater numbers of objects to take up more room, controlling for object size, in addition to the constraint that dots could not overlap on the screen. Alternatively, subjects may have preferred to increase the density (numbers of objects within the enclosing circle) out of an unrelated desire to fill larger spaces with more dots; i.e., they preferred to maintain a particular density level in their responses that was near the maximum density across the experiment.

2.12.4 Effects of brightness

Finally, there was a marginal effect of brightness (opacity) on object size suggesting that brighter objects were remembered as being larger, when controlling for size. The simplest explanation for this is that brighter objects are of greater intensity. If light sources of the same kind tend to be of a homogenous brightness level in the natural world, then bigger (or nearer) objects would tend to be brighter; thus, subjects could misremember brighter objects as being larger (or nearer). However, this does not explain why there would not also be an effect of size on brightness, unless size were estimated only secondarily from early sensory estimates of brightness; answering that question would take further studies.

2.12.5 Conclusion

In summary, the 6 studies yielded a set of idiosyncratic effects that could each warrant their own causal explanation. Given the lack of an overarching pattern, we conclude that, at least in delayed-estimation, dual production tasks do not yield any evidence for a generalized magnitude system. However, the most influential dimension, independently of the presence of patterns, was numerosity: it could be that knowledge of the number of objects in a scene gives rise to particular sensory expectations for object size and brightness that arise in calculating number from static scenes. More generally, the internal models underlying each of the idiosyncratic results could reflect prior knowledge about separate causal processes in the environment; thus, understanding what generates the results seen in Figure 2.1 will require a more thorough understanding of the kinds of causal relationships between each of these pairs of variables in the environment.

These results do *not* indicate that such a generalized magnitude system does not exist: it simply means that if one does exist, it may fail to play any critical role in performance in these tasks because the tasks do not invoke it. One possibility is that a generalized magnitude system has nothing to do with combining or binding together different values from different magnitude systems for a general perceptual problem; instead, it may be something more like meta-knowledge about the scaling of each domain (eg., domain-general representations of relative magnitude), through which different values from different domains can be compared. It may be at this level that subjects experience stronger inter-dimensional interference and bias effects: categorization tasks (also called bisection tasks) that invoke implicit 'larger' or 'smaller' abstract representations may be at this more relevant level. Further studies should therefore compare the biasing effects in bisection tasks to those found here in production tasks: differences in the patterns of influence using the same stimulus values would indicate a crucial role for response type in conclusions about when and if a generalized magnitude system is invoked in dual-task experiments.

Chapter 3. Spontaneous, modality-general abstraction of a ratio scale.

3.1 Preview

The existence of a generalized magnitude system in the human mind and brain, as proposed by Walsh (2003), remains elusive because it has not been clearly defined. In this chapter, we show that one possibility is the representation of relative magnitudes via ratio calculations: ratios are a naturally dimensionless or abstract quantity that could qualify as a common currency for magnitudes measured on vastly different psychophysical scales and in different sensory modalities. In a series of demonstrations based on comparisons of item sequences inspired by literature on melodic-contour representation in music, we demonstrate that subjects use knowledge of inter-item ratios to judge their similarity within and across sensory modalities and across magnitude domains. Moreover, they rate ratio-preserved sequences as more similar to each other than sequences in which only ordinal relations are preserved, indicating that subjects are aware of differences in levels of relative-magnitude information preservation. The ubiquity of this ability across many different magnitude pairs, even those sharing no sensory information, suggests a highly general code that could qualify as a candidate for a generalized magnitude representation.

3.2 Introduction

Spontaneous and flexible mappings among spatial, temporal, and numerical representations suggest the existence of an abstract encoding mechanism that mediates crosstalk between these dimensions. Building on work suggesting a similar format for representations of object length or size, event duration, and numerosity (Gallistel & Gelman, 2000), Walsh (2003) proposed that a domain-general representation of magnitude could explain interference and bias effects in dual magnitude-judgment tasks in human adults. The main sources of evidence for this claim were drawn from disparate literatures: studies from the numerical cognition literature reporting interactions between numerical and various types of spatial representations (location and size) and studies from the literature on the language-cognition interface reporting the influence of the spatial representations (length) on event-duration perception.

For example, when asked to simultaneously monitor a line's length and duration but asked to respond to one dimension only, subjects are typically biased by or experience interference from the magnitude of the unreported stimulus dimension. When judging duration, subjects report that shorter lines last more briefly and report that longer lines last for a greater period of time (eg., Casasanto & Boroditsky, 2008; Merritt, Casasanto, & Brannon, 2010). These results are consistent with competition or interaction between domain-specific inputs at a higher-level of representation. However, as we saw in Chapter 2, inter-dimensional bias effects may be the result of idiosyncratic mechanisms of interaction between dimensions and/or may be highly task-specific.

The term 'generalized magnitude' is not currently well defined. Most studies on this higher-level or abstract representation have focused on imprecise questions about its architecture and at what stage of processing it could arise. For instance, the abstract representation, if it exists, may be implemented either explicitly or implicitly: an explicit implementation would consist of a process in which information from individual magnitudes would be transmitted to a single, shared representation, while an implicit implementation would consist of mappings between dimension- or domain-specific codes containing similar abstract structures. In addition, abstract representations may be generated at early and/or late stages of information processing; this is typically framed as identifying whether magnitude abstraction is a perceptual process independent of verbal or other kinds of culturally specific abstraction or one that belongs solely to task-specific response generation (eg., Chen & Verguts, 2010; van Opstal & Verguts, 2013). Both of these proposals could qualify as a generalized or abstract magnitude representation, but may solve different, coexisting types of computational problems.

Regardless of the specifics of its implementation, the ability to relate dimensions based on abstract magnitude information seems to reflect one or more core conceptual capacities in place long before adulthood. Both infants' and adults' representations of relative magnitude seem to support (1) the binding of stimulus values across sensory modalities in memory and (2) generalization of a concept of 'more' or 'less' across magnitude dimensions. For example, infants bind together relatively short and long durations of various kinds of auditory stimuli (eg., tones or sequences of syllables) with simultaneously presented short and long lines (9-month-olds: Srinivasan & Carey, 2010; newborns: de Hevia, Izard, Coubart, Spelke, & Streri, 2014). Some pairs of dimensions appear to bind together more naturally than others (eg., length and duration are easier to bind than length and loudness), which is consistent with the interpretation that only some between-magnitude mappings are relevant to inferring some causal structure in the world that generates positively correlated magnitudes in sensory data. However, from the data currently available, it is not clear if this variance in ease of stimulus-dimension mapping stems from the problems associated with *binding* causally related representations simultaneously across presented dimensions (or conversely, *segregating* unrelated ones) or from fundamentally different kinds of abstract representations associated with different pairs or groups of dimensions. Answering this question involves resolving the much larger issue of *why* magnitudes might be spontaneously bound together and how that may or may not be connected to the formation of an abstract representation of relative magnitude in the first place. Multiple levels of abstraction—and, more broadly, multiple kinds of interaction between space, time, number, and other domains such as brightness and loudness—are consistent with Walsh's original hypothesis. We come back to this point in the General Discussion, Section 3.21.

A different way of approaching abstract magnitudes is by assessing generalization—i.e, the spontaneous transfer of abstract relations from one dimension to another—rather than binding of simultaneously presented magnitude dimensions. Relatively few studies have taken this approach. In one example, Lourenco & Longo (2010) showed that infants transfer representations of relative magnitude from one domain to another. Specifically, they demonstrated that when infants learned to associate different categorical feature values with large and small object sizes, they expected a similar association between the categorical feature values and large and small numerosities or durations. In another study, de Hevia & Spelke (2010) showed that following exposure to a series of increasing (or decreasing) dot-array numerosities, 8-month-old infants failed to dishabituate to a series of increasing (decreasing) line lengths, but dishabituated to sequences proceeding in the opposite direction. These studies show that infants may possess an abstract or cross-modal concept of *more* and *less* that distinguish pairs or sequences of numerical, spatial, and temporal magnitudes.

Together, dual-task studies in adults as well as habituation studies in infants are consistent with a remapping of each individual dimension onto a common (and potentially a single) abstract code that seems to preserve information about relative magnitude (eg., large vs. small or more vs. less). However, we know comparatively little about the internal structure of these representations of relative magnitude; they are potentially richer than categorical/ordinal re-scaling of stimulus values into representations of "more" or "less". Ratio-dependent discrimination between two sample values from (most) continuous magnitudes and power-law scaling demonstrated by magnitude production tasks (Stevens, 1975) suggest that more detailed information about the scaling of each dimension may be available in abstraction. In addition, relative-magnitude abstractions may extend beyond spatial, temporal, and numerical continua (see Cantlon, Platt, & Brannon, 2009; Bonn & Cantlon, 2012; and Cantlon, 2012 for review).

3.3 Evidence for rich abstraction in magnitude estimation tasks

Ratio representations of relative magnitudes are a potential candidate for a kind of generalized magnitude code that is conceptually independent of the problem of dimension binding or the problem of multiple-cue combination. Ratios (as opposed to differences between absolute values) are naturally *dimensionless*, meaning that no particular physical metric applies: the original physical metrics that generated the numerator and denominator do not matter, except with respect to calculation error inherited from measurement error or uncertainty. Ratio representations of magnitude pairs in different sensory dimensions or modalities may be similar without needing to invoke a common environmental cause, though preservation of ratios across dimensions may often arise in the natural environment (eg., size-invariance of object representations across changes in physical position).

A long tradition within psychophysics demonstrates that mapping of ratio representations of relative magnitude across dimensions is at least possible given explicit instruction. Successes in the use of the technique of magnitude estimation suggest that human adults can explicitly map proportional changes in magnitude from one dimension to any other (Stevens, Mack, & Stevens, 1960; Stevens, 1975; Shepard, 1981; Luce, 1990; 2002). For example, subjects can observe the difference between the sizes of two objects and use that difference to produce an equivalent magnitude change in the loudness of two sounds. Naïve subjects can generate these mappings after receiving brief training with anchor values for each dimension. Stevens' (1975) anecdotal account of the surprising ease of magnitude production tasks suggests that subjects represent ratios of magnitude pairs (or ratios of magnitudes with a memorized anchor value) within each stimulus dimension and that they could access this representation for constructing ad hoc mappings between dimensions. However, it is not known whether this ratio-based representation of relative magnitudes depends on a task-specific, associative mapping between the stimulus anchor points, or whether it is a spontaneously generated, task-independent representation.

Some recent evidence suggests that proportions or fractions in multiple stimulus types such as relative line length or proportions of dots painted in one color in a display with dots of two colors are spontaneously represented in a fronto-parietal network in adult humans and macaques (Vallentin & Neider, 2008; 2010; Jacob & Nieder, 2009; Jacob, Vallentin, & Nieder, 2012), but these results do not specifically predict the degree to which different types of ratio representations can be spontaneously mapped to each other.

3.4 Overview of experiments

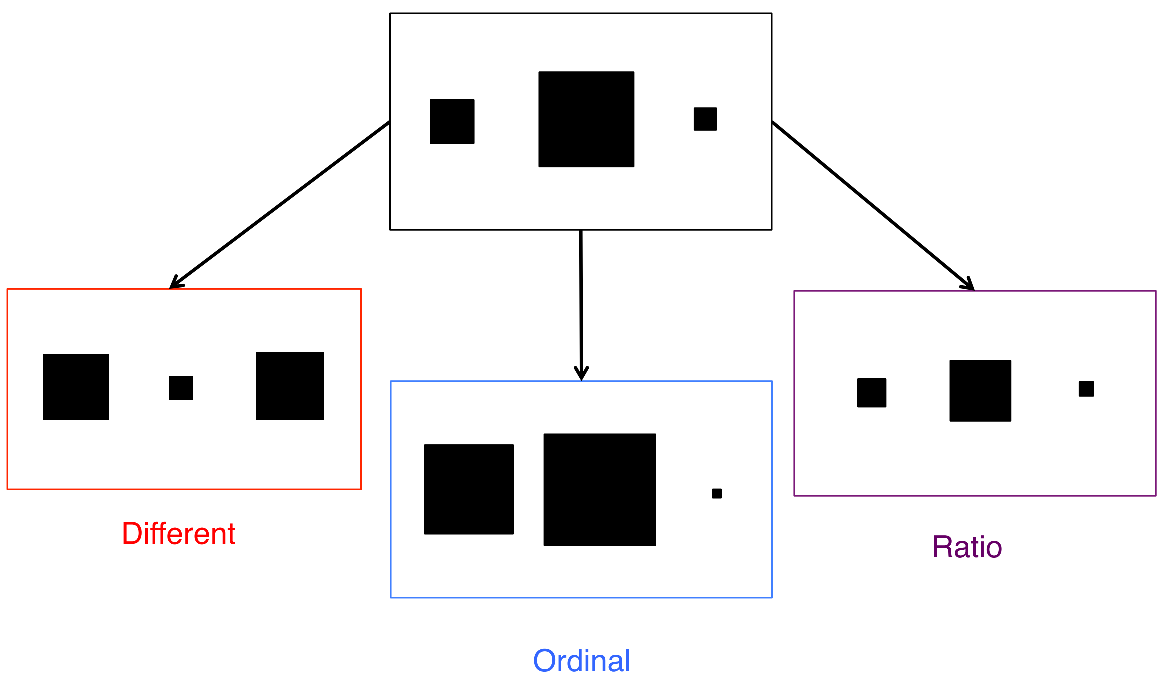
There is substantial evidence for implicit interactions between different magnitudes and there is evidence that people can transfer coarse information about relative magnitude from one dimension into another dimension based on rank ordering, but there is no evidence that people spontaneously represent and use ratio relations to compare values between dimensions, and across modalities. In this paper, we demonstrate that human adults can spontaneously represent magnitude ratios in sequences of visual and auditory objects and use these representations to relate sequences of stimuli across sensory modalities and across dimensions of magnitude.

The music-cognition literature, though seldom cited as evidence for abstract magnitude representations, has long documented humans' abilities to represent relative pitch information, and serves as our main methodological paradigm for probing spontaneous representations of abstract magnitudes. Many studies suggest that human infants and adults as well as macaques retain at least coarse-grained representations of the pitch patterns of melodies in working memory (Dowling & Fujitani, 1971; Trehub, Thorpe, & Morrongiello, 1988; Brosch, Selezneva, Bucks, & Scheich, 2004). The patterns, termed 'melodic contour', are commonly described as the set of pitch-change directions in a series of musical notes, though music-theoretic literature has proposed a more detailed representation of the rank ordering of pitches according to fundamental frequency (Marvin & Laprade, 1987; Marvin, 1997). Most recently, some studies have proposed that these coarse-grained, representations of musical patterns belong not just to musical notes with pitch content associated with periodicity in the acoustic waveform, but within and across other auditory continua such as brightness (correlated with the spectral centroid) and loudness, correlated with amplitude (McDermott, Lehr, & Oxenham, 2008). These results suggest that representations of sets of changes in direction (up/down pitch motions, gains/losses in loudness, etc.) are an auditory-general phenomenon. In addition, one study suggests that adults can relate melodic contour in the auditory domain to line drawings that represent long sequences of up/down pitch changes (Prince, Schmuckler, & Thompson, 2009), suggesting the existence of an abstract, gist-memory mechanism that allows human adults to generalize relative pitch information to the visual modality in addition to across different psychoacoustic continua. To our knowledge, no evidence that representations of relative pitch, loudness, or auditory brightness are special cases of a more general ability to automatically represent sequences of ratios has appeared in the literature.

Using the sequence-pair methodology from the music-cognition and auditory perception literature, we tested the hypotheses that subjects (1) can automatically extract relative-magnitude information at multiple levels of abstraction (ratio and rank) within visual and auditory modalities, (2) can use it to compare sequences that share no sensory information (i.e., it represents an abstraction away from sensory-specific scaling) and (3) and can use this abstract ratio information can be used to compare sequences across space, time and number.

Our basic strategy was to create pairs of 3-item object sequences consisting of a randomly generated standard sequence and comparison sequences that preserved varying levels of abstract structure to the standard. We chose 3-item sequences as the minimum number needed to distinguish rank-ordering representations from simple direction-changes and to impose the absolute minimum working-memory load necessary to detect sensitivity to preserved ratio relationships between items. The comparison could be the same sequence (within-modality and dimension), a sequence in which between-item ratios were preserved, a sequence in which only the rank-ordering of items was preserved, and a sequence in which items were constrained to have different rank-ordering than the standard; see Figure 3.1 for an illustration. Each of these levels of increasing abstraction represents a kind of information loss, so we predicted that perceived similarity of patterns would be a decreasing function of the level of abstraction required to map one pattern onto another: same > ratio-preserved > rank-order preserved > different.

Figure 3.1. Illustration of sequence manipulations. The top box illustrates a sample standard sequence. Time runs from left to right, with each square appearing one at a time. The arrows point to boxes containing the sample manipulations: one constrained to be different (red), one maintaining ordinal but not ratio relationships (blue) and one preserving ratio relationships (purple). The same sequence type is not pictured.



3.5 Experiment 7: Within-dimension sequence comparisons

We asked adults to rate the similarity of pairs of visual and auditory object sequences. The experiment had three, specific objectives: (1) to conceptually replicate and extend the findings of McDermott et al. (2008) in the auditory domain using continuous similarity ratings, (2) to demonstrate pattern-memory effects in visual event sequences analogous to those found in auditory event sequences, and (3) to test whether humans are sensitive to manipulations of specific kinds of mathematical relations among stimulus magnitudes.

Visual sequences consisted of three squares varying in either height—location along the vertical midline—or area. Auditory sequences consisted of three, band-pass-filtered, white noise samples varying in either height (center frequency) or loudness (a combination of bandwidth and volume). Each trial consisted of two sequences: the standard (first) and comparison (second). Comparison sequences were either identical to the standard (*Same* sequences), similar with inter-item ratios preserved (*Ratio* sequences), similar with only inter-item ranks preserved (*Ordinal* sequences), or dissimilar sequences generated pseudorandomly (*Different* sequences).

We predicted that similarity ratings for *Same* sequences would be highest, while those for *Different* sequences would be lowest. Ratings for the *Ratio* and *Ordinal* sequences would be between these extremes. Critically, we predicted that *Ratio* ratings would be higher than *Ordinal* ratings if subjects were sensitive to the perturbation or loss of ratio information.

3.6 Method

3.6.1 Subjects

Adults were recruited via Amazon Mechanical Turk (www.mturk.com) and directed to an external website; n = 15 for each of the four within-dimension conditions. Subjects had to possess a 95% approval rating and be located within the United States to view and accept the task. Upon satisfactory completion of the task, subjects were paid a fixed amount of $3.50 (approximately $8/hr).

Sample sizes for each condition were determined on the basis of inspection of the individual results of 3 (independent) pilot subjects in the Object Size condition. We collected data with the goal of at least 20 data points per cell for each of 15 subjects per condition to estimate multilevel linear regression models with maximal random-effects structures.

3.6.2 Stimulus Generation

Testing required participants to navigate to an externally hosted website constructed with custom software. In addition to the standard HTML/CSS markup languages, all sequences were constructed using a combination of standard Javascript as well as the JQuery (www.jquery.com) and JQueryUI (www.jqueryui.com) libraries. Auditory stimuli were rendered in the browser via the WebAudio API, available in recent versions of Google Chrome, Opera, and Mozilla Firefox[1]. Visual stimuli were rendered in HTML5's Canvas utility with the fabricjs library (Zaytsev & Chernyak, 2014). Experiments without sound could be completed in recent versions of Safari as well as Chrome and Firefox.

3.6.3 Stimuli

There were 4 sequence types: 2 within-vision (height or size) and 2 within-auditory (noise brightness or noise loudness). We chose these dimensions based on the non-pitch experiments found in McDermott et al. (2008) and intuitively related visual analogues.

Each standard sequence was paired with one of four, possible comparison stimuli: it was either (1) the same as the first (in within-modality manipulations only), (2) an affine transformation of the first such that the inter-item ratio (and magnitude rank) present in the first stimulus was preserved but the absolute values were not, (3) a transformation in which the ordinal (rank) relations among the items in the first sequence were preserved but not the inter-item ratios, and (4) a pseudorandom sequence in which both the absolute values and inter-item relations were constrained to be uncorrelated with the first sequence. More specifically, the Pearson correlation coefficient of the absolute values was constrained to be between -0.1 and 0.1 and at least one of the between-item difference values had to change sign. For example, if a visual stimulus increased in size from position one to position 2 and also from position 2 to 3 in the standard sequence, the comparison sequence stimuli had to include one size change in which the stimulus decreased in size. Stimulus values for each trial were generated offline in the R statistical computing package.

Visual stimuli were enclosed in a white, 600-by-600-pixel HTML Canvas frame with a black, 1-pixel-thick border. Each sequence consisted of three, 500-ms stimuli separated by 250ms inter-stimulus intervals. In all sequence, successive-stimulus magnitude ratios were constrained to be no smaller than 7:8 when comparing the smaller to the larger value. In the object-size sequences, squares were constrained to be from 50 to 500 pixels in width (and height). In the object-height sequences, they were constrained to be located within a 500-pixel window with invisible, 100-pixel borders at the top and bottom.

Auditory stimuli were presented with a similar visual layout, though during sound presentation an icon appeared and stayed in the center of the Canvas window. Each sequence consisted of three, 500-ms stimuli separated by 250-ms inter-stimulus intervals. All sounds consisted of a single sample of white noise generated offline but each sound was filtered in real time in the web browser. Each sound consisted of 5-ms ramp-up and ramp-down periods to minimize the effects of auditory transients.

In noise-brightness sequences, for each sequence, subjects heard 3 sounds in which the center frequency, but not the proportional width, of the filter, was varied. The frequencies were constrained to be between 1000Hz and 10000Hz and scaled in equivalent-rectangular-bandwidth units to ensure appropriate psychophysical scaling. In the loudness condition, subjects heard 3 sounds in which the width of the filter and the amplitude, but not the center frequency, was varied. The simultaneous filter-width and amplitude manipulations were designed to allow for a wide range of differences in apparent loudness without imposing potentially uncomfortable stimulus levels, given that the absolute volume level for individual subjects is impossible to directly control on Mechanical Turk. Subjects were urged in practice trials involving auditory stimuli to adjust their volume to comfortable levels; to ensure maximal comfort, the volume range of these stimuli was designed to be representative of the range of possible values.

3.6.4 Procedure

After accepting the task on the Mechanical Turk website, subjects read a set of general instructions including a request to close all other tabs and windows to maximize browser performance. They then filled out an optional demographics form. Finally, they read specific instructions about the sequences they would encounter. They were asked to pay attention to the patterns of size, height, pitch, or loudness without any instruction regarding what kind of pattern needed to be extracted, as is standard in melodic contour experiments (Dowling & Fujitani, 1971).

On each trial, the onset of each sequence was preceded by the words 'Sequence 1' or 'Sequence 2' printed in the middle of the viewing area. These lasted for 750 ms each. At the end of the trial, subjects had to adjust a vertical slider that was 400 pixels tall on the right side of the screen. They were then asked to rate the similarity of the sequences using a 400-pixels-tall slider next to the Canvas. They were prompted to do this with the words 'Enter Response,' after which they clicked the 'Next Trial' button to proceed. The lower end and upper ends of the slider were labeled with the texts ‘Very Different’ and ‘Very Similar,’ respectively. The slider handle always began in the middle of the slider at 200px. After moving the slider, the subjects were instructed to record their final response by pressing a button with the mouse to move on to the next trial.

The full experiment consisted of two practice trials (of type *Same* and *Different* with no feedback), then 20 trials of each pair type at test, yielding 82 trials per subject. Subjects proceeded from practice trials to test trials without interruption. Trials were presented in random order using an implementation of the Fisher-Yates algorithm.

To ensure subjects were paying attention, on 5 supplemental catch trials in visual and auditory conditions, the visual stimulus was replaced with the words ‘CATCH TRIAL.’ On these trials, subjects were required to report that they had seen those words by pressing a button labeled ‘Catch Trial.’

Following completion of the experiment, subjects submitted their data to the server and were given a randomly generated hexadecimal code to submit to Mechanical Turk for payment.

3.6.5 Exclusion of Subjects

Because we could not directly control the performance of subjects’ hardware or web browsers, we expected the need to exclude subjects who performed seemingly at random independently of performance on catch trials. We excluded subjects whose responses in the ‘same’ condition were not significantly above the midpoint (the 200-pixel halfway point on the slider) in independent, one-tailed, one-sample t-tests (88% of subjects). This amounts to a test of sequence recognition accuracy. Independent replication of the results with a different set of stimulus values verified that our results were not dependent on these subject-exclusion strategies or particular stimulus sets.

3.7 Results

3.7.1 Analytical Approach

We estimated effects using linear, multilevel models (Gelman & Hill, 2007) fitted using the restricted maximum likelihood estimator. We included the maximal set of random effects (random intercepts and slopes by subject and random intercepts by stimulus) in each model. Significance of the resulting t-values was assessed using Kenward-Roger (Kenward & Roger, 1997) approximate degrees of freedom.

Given the bounded nature of all ratings measures, ratings within each condition were distributed non-normally and heteroskedastically, with compression near the upper and lower limits of the range. To correct for this, we rescaled the dependent measure: first, we changed the raw measure in pixels (from 0 to 400) to be between 0 and 1 by adding 1 and dividing by 402. We then took the logit transform of the rescaled responses, where $logit(x) = ln\ x/(1-x)$.

We coded the conditions using backward-difference codes, which assigned each successive pair of levels of a categorical predictor corresponding to their mean differences. Each coefficient thus represents the mean difference (in logit space) in the transformed rating between the following pairs of conditions: *Different* vs. *Ordinal*, *Ordinal* vs. *Ratio*, and *Ratio* vs. *Same*.

3.7.2 Regression results

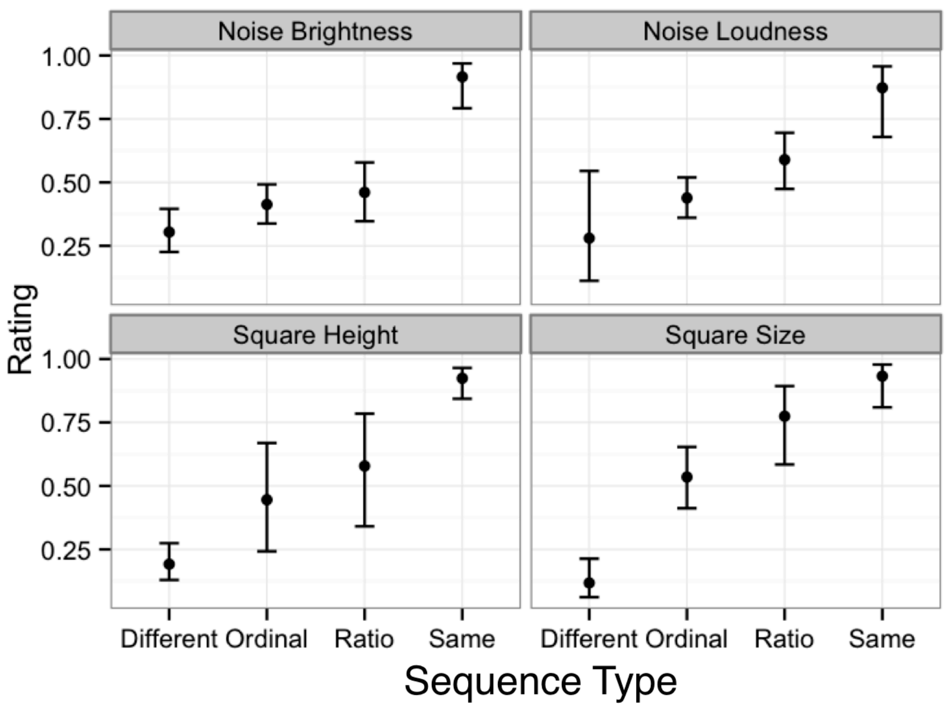
We first analyzed each condition separately, as the purpose of the experiment was to detect differences in rating between sequence types rather than to compare the strength of the ratings differences across stimulus dimensions. All modeling results are presented in Table 3.1. The mean ratings (scaled from 0 to 1, with 0 being most dissimilar) with 95% confidence intervals are presented in Figure 3.2.

Table 3.1. Regression results for Experiment 7: within-dimension sequence pairs. Degrees of freedom for the *t*-statistics (in the leftmost column) reflect Kenward-Roger approximations. Coefficients for condition effects represent differences in condition means in logit-transformed ratings.

Condition	Coefficient	*B*	*SE*	*t*	*p*
Object Height	Intercept	0.29	0.3	0.97	0.3521
$n = 12$	Different vs. Ordinal	1.22	0.54	2.26	0.0448
$df \approx 11.00$	Ordinal vs. Ratio	0.54	0.14	3.93	0.0024
	Ratio vs. Same	2.18	0.46	4.78	0.0006
Object Size	Intercept	0.5	0.2	2.45	0.0263
n = 14	Different vs. Ordinal	2.15	0.44	4.92	0.0002
$df \approx 15.74$	Ordinal vs. Ratio	1.09	0.31	3.57	0.0026
	Ratio vs. Same	1.39	0.33	4.22	0.0007
Noise Brightness	Intercept	0.26	0.1	2.58	0.0197
$n = 14$	Different vs. Ordinal	0.48	0.27	1.75	0.0983
$df \approx 16.84$	Ordinal vs. Ratio	0.19	0.3	0.65	0.5233
	Ratio vs. Same	2.54	0.64	4	0.001
Noise Loudness	Intercept	0.27	0.11	2.6	0.0216
$n = 13$	Different vs. Ordinal	0.7	0.32	2.19	0.0467
$df \approx 13.21$	Ordinal vs. Ratio	0.61	0.23	2.69	0.0185
	Ratio vs. Same	1.56	0.67	2.33	0.0365

Note that the coefficients for nearly all comparisons are significantly different from zero, with two exceptions in the noise-brightness condition, the *Different* vs. *Ordinal* comparison was marginally significant, while the *Ordinal* vs. *Ratio* comparison was not. However, the pattern of results in this condition was qualitatively in the same direction as the rest of the experiment. See the follow-up Experiment 8 for a confirmation of the robustness of this pattern.

Figure 3.2. Mean ratings and 95% confidence intervals for Exp. 7.



3.7.3 Aggregate Analysis

We ran a post-hoc analysis of all conditions in Experiment 7 to explore whether or not the noise-brightness condition differed significantly from the others. The aggregate model was constructed in the same way as the individual models, except: (1) the additional factor of 'Condition' was entered with a simple coding scheme (centered dummy codes for minimization of nonessential multicollinearity); (2) the random intercept by individual stimulus removed, and (3) *p*-values used Satterthwaite's approximation for efficient computation (Satterthwaite, 1946); the noise-brightness condition was coded as the reference group. Results are shown in Table 3.2.

The aggregate analysis revealed the expected main-effect differences by pair type, indicating that overall subjects were sensitive to the sequence manipulations. The analysis also revealed a main effect of condition, indicating that average ratings in the task varied by individual condition. The coefficients for each simple code revealed significant differences between the average ratings in the noise-brightness condition and both visual conditions (size and height), but failed to reveal a difference in average ratings between the noise-brightness condition and the loudness condition. Thus, noise-brightness ratings were lower overall when compared with the visual conditions, which is not surprising given the low ratings for everything but the *Same* sequence types in that condition.

The set of interaction terms revealed each sequence type comparison varied by each condition contrast, meaning that the magnitude of the differences between ratings in successive pairs of sequence types (different vs. ordinal, ordinal vs. ratio, and ratio vs. same) was different between the noise-brightness and other conditions.

Table 3.2. Aggregate analysis of Experiment 7 results. The set of coefficients for sequence-pair type are backward-difference coded as in the separate analyses for each condition and condition is simple-coded with noise-brightness as the reference group. The interaction terms are abbreviated using the first letter of the main-effect conditions: for example, Different vs. Ordinal becomes DvO in the set of interaction terms. Degrees of freedom from the ANOVA-summary column are uncorrected (not necessary here due to very large ratios) but individual predictors are corrected using the Satterthwaite approximation; the Kenward-Roger approximation is much less efficient for large models with many random effects.

ANOVA summary	Coefficient	*B*	*SE*	*df**	*t*	*p*
	(Intercept)	0.32	0.16	32.90	1.94	0.0615
F(50.71, 3)*	Different vs. Ordinal	0.98	0.31	31.80	3.22	0.0030
	Ordinal vs. Ratio	0.58	0.10	18.10	5.77	0.0000
	Ratio vs. Same	1.84	0.18	52.00	10.06	0.0000
F(24.54, 3)*	Brightness vs. Height	2.69	0.24	244.80	11.46	0.0000
	Brightness vs. Loudness	0.09	0.18	415.20	0.51	0.6092
	Brightness vs. Size	1.19	0.39	37.90	3.00	0.0047
F(27.16, 9)*	DvO:BvH	5.24	0.44	149.50	11.86	0.0000
	OvR:BvH	1.56	0.20	105.20	7.81	0.0000
	RvS:BvH	-1.01	0.42	109.70	-2.42	0.0172
	DvO:BvL	0.29	0.35	199.90	0.81	0.4171
	OvR:BvL	0.46	0.19	224.20	2.41	0.0167
	RvS:BvL	-0.79	0.36	176.90	-2.20	0.0291
	DvO:BvS	3.23	0.74	36.40	4.39	0.0001
	OvR:BvS	1.30	0.25	24.00	5.13	0.0000
	RvS:BvS	-1.12	0.47	61.70	-2.36	0.0214

The exception was a failure to find a difference between the magnitude of the difference in *Different* and *Ordinal* ratings between the noise-brightness and noise-loudness conditions. This indicated that the noise-brightness condition not only differed in average ratings, but also differed in the average rating differences between sequence types.

3.8 Interim discussion

Overall, the results from Experiment 7 suggest that subjects can detect differences in levels of information preservation across standard and comparison sequences in the same stimulus dimension and modality. Specifically, and most importantly, subjects rated sequence pairs in the *Ratio* sequence types as more similar than those in the *Ordinal* sequence types in 3 of the 4 conditions.

In addition, the results in the visual domain suggest that representations of (fine-grained) information about relative magnitudes in sequence-comparison tasks are not simply a property of the auditory or music-cognition system but one that is potentially more general with respect to magnitudes in different sensory modalities.

In the noise-brightness condition the analysis failed to detect subjects' sensitivity to any similarities between sequences except in *Same* sequences; the aggregate analysis revealed significant differences in performance in the average ratings and differences between ratings of each sequence-pair type. However, because the results showed the same general pattern as the other 3 conditions and the coefficients for each sequence-pair contrast had the correct sign, subjects may have failed because they tended to rely more on absolute frequency than relative frequency in this task. Subjects would thus have subjectively divided the rating space into 'Same' and 'Not-same' categories, compressing the differences between the 'not-same' sequence pairs. We explored this possibility with a follow-up experiment in which we removed the 'Same' sequence type and increased the number of trials to increase power.

3.9 Experiment 8: Noise-brightness, revisited

In the noise-brightness condition, the estimated differences between *Ratio* and *Ordinal* and between *Ordinal* and *Different* sequence pairs were not significant, but in the correct direction or pattern. There are several possible reasons for this failure to detect differences. One is a simple lack of power. Another is that subjects possibly over-weighted absolute pitch cues, resulting in a large difference between the *Same* pairs and the other three pair types and thus compressing the ratings for the other three conditions.

To test this secondary hypothesis, we ran an additional noise-brightness condition and eliminated the *Same* condition to maximize the subjective weighting of relative over absolute pitch (center-frequency) cues.

3.10 Method

We used the same methods as detailed in Experiment 7 for the noise-brightness condition, with the following exceptions. First, we eliminated the *Same* sequence type and increased the number of trials in each of the remaining sequence types from 20 to 30. Second, our original test of recognition accuracy was not available (due to the lack of a 'Same' condition). We were mainly interested in the difference between the 'Ordinal' and 'Ratio' conditions, so we excluded subjects based on whether or not each subject rated 'Different' sequences significantly below the slider midpoint in independent, one-tailed, one-sample t-tests. Independent replication of the results with a different set of stimulus values verified that our results were not dependent on these subject-exclusion strategies or particular stimulus sets.

3.11 Results

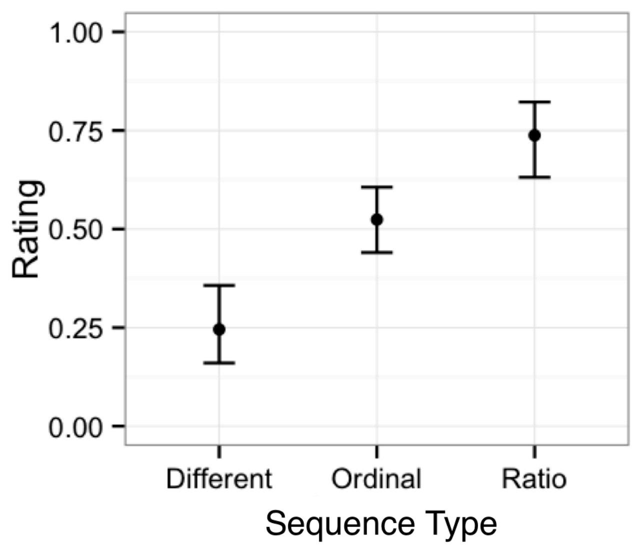
We followed the same analytical approach as in Experiment 7, with the following exception: we dummy coded the sequence-type variable, with the ordinal sequence type as the reference group. This gives equivalent results and is the simpler alternative to backward difference coding available with only 3 groups and 2 pairwise comparisons. These results are presented in Table 3.3 and in Figure 3.3.

Table 3.3. Experiment 8 results.

Condition	Coefficient	*B*	*SE*	*t*	*p*
Pitch Height (No Same)	Intercept	0.1	0.17	0.56	0.5834
$n = 14$	Ordinal vs. Different	-1.22	0.34	-3.56	0.0026
$df \approx 16.11$	Ordinal vs. Ratio	0.94	0.25	3.77	0.0017

Ratings in this follow-up experiment showed significant differences between the *Different* and *Ordinal* conditions as well as between the *Ordinal* and *Ratio* conditions. Specifically, the *Different* sequence types were rated lower than the *Ordinal* sequence types and the *Ratio* sequence types were rated higher than the *Ordinal* sequence types.

Figure 3.3. Mean ratings and 95% confidence intervals for Exp. 8.



3.12 Interim Discussion

The results were consistent with our hypothesis that subjects had weighted absolute center-frequency cues very highly when *Same* trials were present in the original stimulus set. Removal of these trials not only allowed time for more trials to be added to increase power, but subjects spread out their responses to cover a fuller range of the slider rating space. For example, in this experiment, the *Ratio* sequence type had a mean rating near 0.75, whereas the mean rating for *Ratio* sequence types in Experiment 7 was just below 0.5; in addition, their 95% confidence intervals did not overlap, indicating a significant difference in ratings for *Ratio* sequence types between experiments.

Most importantly, this change in the amount of separation between ratings for different sequence types revealed that subjects were sensitive to the difference between *Ratio* and *Ordinal* sequence types in the domain of center frequency (auditory brightness).

The reliance on absolute pitch cues in the subset of results from Experiments 7 and 8 is somewhat surprising, as auditory brightness is closely related to perception of pitch height: usually human infants and adults pay attention to *relative* pitch cues in melody recognition (eg., Plantinga & Trainor, 2005). However, this finding is consistent with an emerging literature that shows that human adults do pay attention to absolute pitch cues (Levitin & Rogers, 2005). It may also be the case that relative pitch cues are more salient for highly periodic, musical tones than for noises in which only the center frequency of the filter is the correlate for the pitch-height dimension.

3.13 Experiment 9: Cross-modal sequence comparisons

In Experiments 7 and 8 we showed that ratio-abstraction supports similarity ratings within the same magnitude dimensions. However, a domain-general code, whether shared as a common resource or simply a common code generated by all systems representing magnitudes, should be able to support comparisons across domains that share no sensory information. In Experiment 9, we demonstrate that fine-grained representations of relative magnitude can be compared across the visual and auditory modalities.

We used the dimensions from Experiments 7 and 8 and asked subjects to compare object height to noise brightness and to compare object size to noise loudness.

3.14. Method

3.14.1 Subjects

Adults were recruited via Mechanical Turk and directed to an external website; $n = 12$ for each of the four cross-dimension conditions. Subjects had to possess a 95% approval rating and be located within the United States to view and accept the task. Subjects were paid $4 due to slightly increased experiment length compared with the first experiment (approximately $8/hr).

We decreased the sample sizes to 12 subjects and increased the number of trials for each of the 3 sequence types to 30 (as in Experiment 8).

3.14.4 Stimuli

Stimuli were generated in the same manner as in Experiments 7 and 8.

There were 4, between-subject stimulus conditions: object-height to pitch-height, and object-size to loudness, and their opposite comparison orders.

For each sequence type, there were three types of relational manipulations performed on the second sequence in each pair: the same *Ratio*, *Ordinal*, and *Different* pairs as experienced in Experiment 7, and generated in the same way. No *Same* condition was possible. However, stimulus values for the comparison sequences were generated in the scale of the standard, then linearly translated to the comparison scales using the range of values explained in Experiment 7, section 3.7.1.

3.14.5 Procedure

The procedure was the same as in Experiment 7, though subjects heard or saw a first standard sequence and then heard or saw a comparison sequence from the other modality. The full experiment consisted of two practice trials (of type 1 and 4 with no feedback), 30 trials of each pair type at test, as well as 5 catch trials, yielding 97 total trials per subject.

3.15 Results

3.15.1 Separate analyses for each condition

We followed the same analytical approach as in Experiment 8, including the same subject exclusion criteria. Regression results for each individual condition are presented in Table 3.4 and means are displayed in Figure 3.4.

All between-sequence-type contrast coefficients were significant, indicating that subjects differentiated between *Ratio* and *Ordinal* sequence types as well as between *Ordinal* and *Different* sequence types.

3.15.2 Aggregate analysis across conditions

As with Experiment 7, we performed an aggregate analysis across all conditions. To reduce nonessential multicollinearity, we centered the dummy-coded sequence types to form simple codes, again with the *Ordinal* condition as the reference group. For each condition, we were interested in whether performance varied by magnitude type (metathetic/qualitative vs. prothetic/quantitative). The distinction between magnitude types is one that

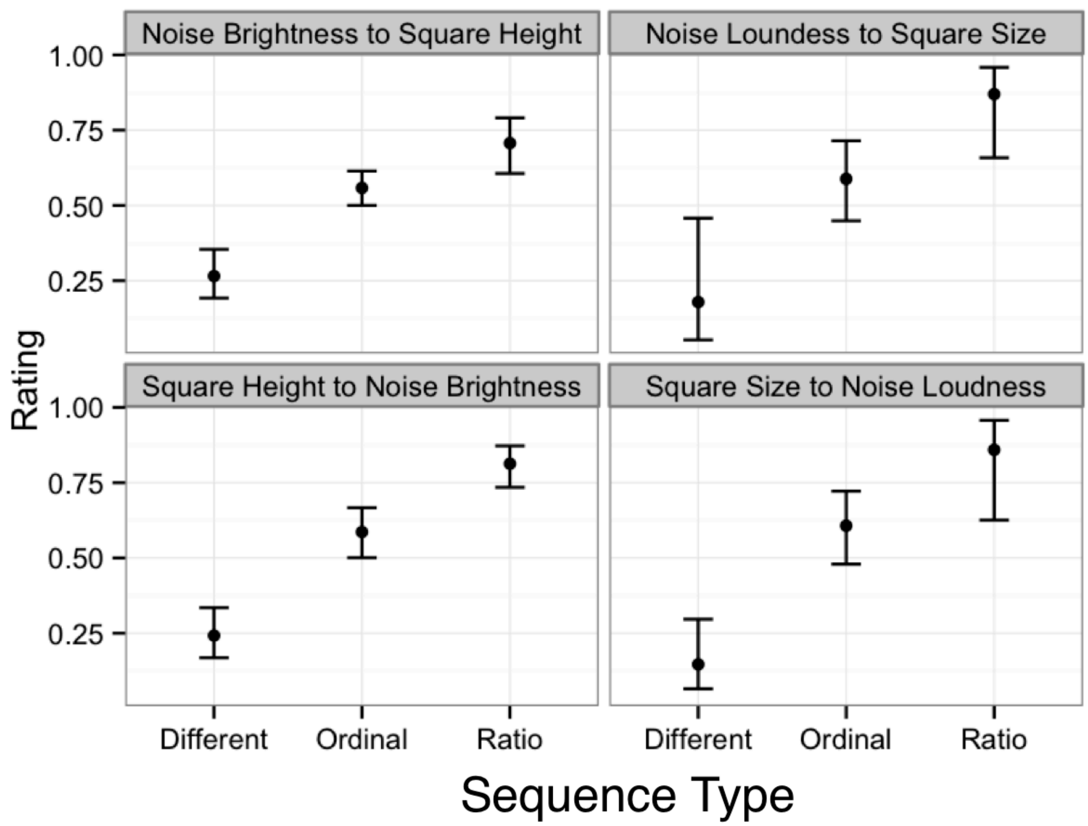
Stevens (1975) made and one that has been proposed as a possible taxonomic distinction related to magnitude dimensions that could form of a generalized magnitude system: prothetic magnitudes might be connected by a generalized magnitude system because they can be characterized as amounts, whereas metathetic magnitudes might not because their description is specific to the unique properties of each domain (Lourenco & Longo, 2011). In addition, we wished to account for any possible order effects.

Therefore, we used a contrast-coding scheme in which (1) the object-height and noise-brightness conditions were compared with the object-size and noise-loudness conditions (the Metathetic vs. Prothetic contrast), (2) the noise-first conditions were compared with the noise-last conditions, and (3) a third, balancing contrast that completed the set. Results are displayed in Table 3.5.

Table 3.4. Experiment 9 results.

Condition	Coefficient	*B*	*SE*	*t*	*p*
Loudness to Size	Intercept	0.36	0.29	1.24	0.2437
n = 10	Ordinal vs. Different	-1.88	0.83	-2.25	0.0495
df ≈ 9.38	Ordinal vs. Ratio	1.54	0.52	2.94	0.0157
Size to Loudness	Intercept	0.44	0.26	1.65	0.1257
n = 12	Ordinal vs. Different	-2.2	0.57	-3.86	0.0023
df ≈ 11.95	Ordinal vs. Ratio	1.37	0.53	2.58	0.0243
Object Height to Noise Brightness	Intercept	0.35	0.18	1.98	0.0639
n = 8	Ordinal vs. Different	-1.49	0.23	-6.35	0
df ≈ 17.33	Ordinal vs. Ratio	1.12	0.27	4.14	0.0007
Noise Brightness to Object Height	Intercept	0.23	0.12	1.96	0.0905
n = 7	Ordinal vs. Different	-1.25	0.27	-4.57	0.0026
df ≈ 7.00	Ordinal vs. Ratio	0.65	0.19	3.47	0.0104

Figure 3.4. Mean ratings and 95% confidence intervals for Exp. 9. (No *Same* condition.)



This analysis yielded the expected main-effect differences in mean ratings by sequence type. However, no main effect of contrast-coded condition was found, indicating no difference in mean ratings for the metathetic/prothetic distinction nor an effect of modality order.

There were, however, two significant interaction terms: the differences in ratings between sequence types varied by magnitude type. That is, ratings were more compressed for the metathetic continua than they were for the prothetic continua. In other words, *Different* stimuli were perceived as more different and

Ratio stimuli were perceived as more similar in the prothetic (size, loudness) continua than in the metathetic (height, brightness) continua.

Table 3.5. Aggregate analysis of Experiment 9. As with the aggregate analysis of Experiment 7, the degrees of freedom for each individual coefficient are estimated using the Satterthwaite approximation. The significance of the *F* ratios for the batches of coefficients are here also estimated with the Satterthwaite approximation. The abbreviation scheme for the interaction terms is as before, using the first letter of each word: Ordinal v. Different changes to OvD and Auditory-First vs. Visual-First changes to AFvVF.

ANOVA summary	Coefficient	*B*	*SE*	*df*	*t*	*p*
	(Intercept)	0.23	0.06	40.03	3.71	0.0006
F(2,70.51)	Ordinal vs. Different	-1.61	0.17	40.33	-9.33	0.0000
= 61.88	Ordinal vs. Ratio	1.09	0.10	44.69	10.56	0.0000
F(3,40.03)	Metathetic vs. Prothetic	0.11	0.25	40.03	0.42	0.6775
= 0.3	Auditory-First vs. Visual-First	-0.11	0.25	40.03	-0.44	0.6648
	Balancing Contrast	0.20	0.25	40.03	0.81	0.4235
F(6,70.51)	OvD:MvP	-2.58	0.69	40.33	-3.72	0.0006
= 4.018	OvR:MvP	1.86	0.41	44.69	4.48	0.0001
	OvD:AFvVF	-0.30	0.69	40.33	-0.44	0.6651
	OvR:AFvVF	0.15	0.41	44.69	0.37	0.7152
	OvD:BC	0.19	0.69	40.33	0.27	0.7884
	OvR:BC	0.42	0.41	44.69	1.02	0.3142

3.16 Interim Discussion

The results of Experiment 9 indicate that subjects can use ratio (and ordinal) representations of sets of magnitudes to compare sequences across sensory modalities, regardless of the order of presentation and regardless of the particular dimension pairs used.

This result builds on the findings of Prince et al. (2009) and McDermott et al. (2008) to show that the concept of 'melodic contour' or up/down pitch-change patterns may reflect a more general capacity for relating relative magnitudes across modalities. Moreover, we have shown that this abstract representation is more fine-grained than previously demonstrated: subjects can use inter-item ratios as well as ordinal relations to compare sequences.

Though there are no large distinctions between the results in the metathetic and prothetic continua that would warrant evidence for a major taxonomic division, there appear to be subtle differences in the pattern of ratings between magnitude types. The origins of this difference are unclear. On the one hand, metathetic continua could be less conducive to producing similar ratio representations between individual stimulus values because the original, subjective measurement scales are different; object height ratios would only map to noise-brightness ratios if the original scales are (at least approximately) similar, or aligned or calibrated in some other way. On the other hand, it could be the case that subjects use a more abstract strategy in these domains: instead of calculating ratios over individual stimulus values, subjects could be calculating 2 *distances* between successive stimulus values in their respective scales by calculating a signed difference value, then calculating a single ratio between the two distances. Further work will be necessary to determine the particular strategy subjects may be relying on.

In addition, this second possibility highlights the fact that prothetic continua (quantitative measurements of amounts) can be created from distance calculations in metathetic continua. Thus, at higher levels of abstraction—when dealing with relations between multiple values from the same domain—the distinction between the two kinds of magnitudes may begin to blur.

3.17 Experiment 10: Relative magnitudes in space, time and number

The previous 3 experiments demonstrated that subjects extract ratio-scaled and ordinal-scaled representations of object sequences and can compare those representations within dimensions and across modalities. However, so far this demonstration has been limited to representations of object height, auditory brightness, object size, and loudness. In this final experiment, we expand on our previous findings to show that similar cross-dimension comparison behavior extends to 2 other canonical magnitude domains as well as object size: time (interval duration) and number (Arabic numerals). These are the three domains identified by Walsh (2003) as the main components of the generalized magnitude system. Thus, demonstration of cross-dimension mappings in these domains would indicate that abstract ratio information could be a candidate for a domain-general magnitude code.

Three-item sequences in this experiment could be squares of different sizes, as in Experiments 7 and 9, tones synthesized in the browser paired with a concurrent visual stimulus at varying durations, or Arabic numerals.

3.18 Method

3.18.1 Subjects

We recruited 20 subjects for each pair of magnitude dimensions, 10 for each order of dimension (eg., 10 for number-size and 10 for size-number). Because the size-duration condition took too long to complete, for the number-duration condition we recruited 30 subjects for those pairs, with 15 experiencing each order, and reduced the number of within-sequence-type trials. All other recruitment specifications and payment procedures were identical to Experiments 8 and 9.

3.18.2 Sample Sizes

Sample sizes (within subject) for each sequence type were similar to Experiments 8 and 9—25 for each, adding to 75 test trials, 2 practice trials, and 5 catch trials. In the number-duration condition, we decreased the number of trials per sequence type to 12.

3.18.3 Stimuli

Stimuli were generated using the same browser software and libraries as Experiments 7-9 and stimulus presentation occurred in the same interface as in Experiments 7 to 9.

Number. Numbers were Arabic numerals ranging from 5 to 50 selected randomly from a uniform distribution on a logarithmic scale and rounded to the nearest integer. They were presented in a bluish color (hexadecimal code #0066FF) in the center of the screen for 500 ms each.

Duration. Duration stimulation was conducted in both the auditory and visual modalities to help subjects maintain attention across the experiment; informal piloting in a separate group of subjects showed that bimodal stimulation helped improve performance. Duration stimulation necessarily increases working memory demands, as trials vary in overall length as a function of duration, and due to the reduced numbers of trials available, we opted to give subjects as much help in remembering duration as possible.

Duration stimuli consisted of a red dot with a 25-pixel radius centered both vertically and horizontally in the viewing area. Subjects also simultaneously heard a sinusoidal tone of 440 Hz synthesized in the browser using the WebAudio API), with 5-ms ramp-up and ramp-down times to reduce transients.

Intervals were restricted to last between 500 and 5000 ms and were randomly selected from a uniform distribution on a logarithmic scale (or generated via an affine transform from integer space in number-duration comparisons).

Size. Size stimuli were created with the same restrictions used in Experiments 7 and 9.

All other stimulus restrictions remained the same as in previous experiments.

3.18.5 Procedure

The procedure was the same as in Experiments 7-9.

3.18.6. Exclusion of Subjects.

We anticipated excluding subjects with the same procedure used in Experiments 8 and 9 for the size-number and size-duration experiments, which had enough trials per person to determine whether a subject was not performing the task correctly. However, there was no need to exclude any subjects for these two conditions.

We decided *a priori* to include all subjects for the number-duration experiment, as there were too few trials per person in the *Different* condition to reliably detect sub-par performance.

3.19 Results

3.19.1 Within-condition analyses

As with the previous experiments, we first conducted within-condition analyses using the same dummy-coding scheme as Experiments 8 and 9. However, on the basis of the lack of dimension order effects in Experiment 9, we treated dimension order as a counterbalancing factor and excluded it from the analysis. These results are shown in Table 3.6 and Figure 3.5.

Results from experiment show that subjects rated *Different*, *Ordinal*, and *Ratio* sequence types significantly differently from one another in all conditions.

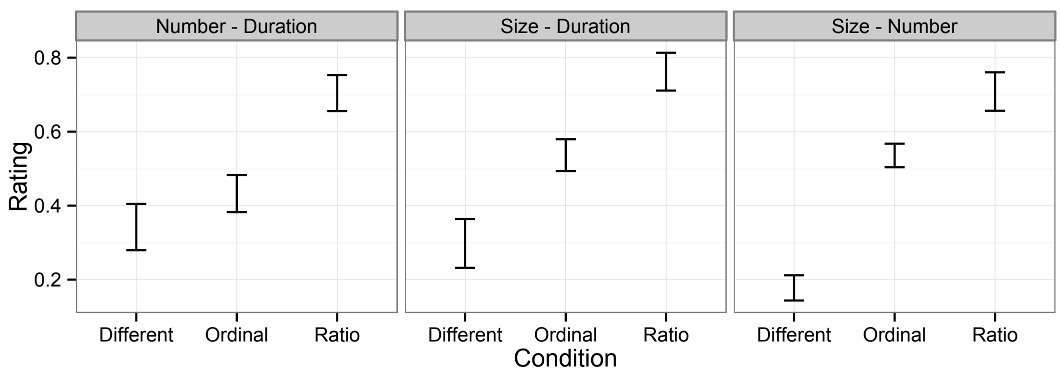
3.19.2 Aggregate Analysis

As with the previous experiments, we conducted an aggregate analysis across conditions. We centered the sequence-pair type, dummy-coded variable to create simple codes. We also simple-coded condition; because performance in the number-duration condition appeared different from the other two, we coded it as the reference group. These results are shown in Table 3.7.

Table 3.6. Experiment 10 results. Degrees of freedom calculated with the Kenward-Roger approximation.

Condition	Coefficient	*B*	*SE*	*t*	*p*
Number-Duration	(Intercept)	-0.27	0.21	-1.33	0.1903
$n = 30$	Ordinal vs. Different	-0.67	0.28	-2.38	0.0228
$df \approx 36.52$	Ordinal vs. Ratio	0.88	0.24	3.74	0.0006
Size-Duration	(Intercept)	0.15	0.17	0.85	0.4023
$n = 20$	Ordinal vs. Different	-0.88	0.32	-2.75	0.0112
$df \approx 24$	Ordinal vs. Ratio	1.19	0.29	4.16	0.0004
Size-Number	(Intercept)	0.14	0.13	1.13	0.2729
$n = 20$	Ordinal vs. Different	-1.55	0.24	-6.56	0.0000
$df \approx 21.52$	Ordinal vs. Ratio	0.90	0.25	3.55	0.0019

Figure 3.5. Mean ratings and 95% confidence intervals for ratings for Exp. 10.



The results show the expected effects of sequence-pair type. They also reveal an effect of condition, with both conditions involving size having higher average ratings than the number-duration condition. This seems to be largely due to the lower ratings for the *Ordinal* sequence type in the number-duration condition. In addition, there was significant interaction for the *Ordinal*-vs.-*Different* contrast and the *Number-Duration*-vs.-*Size-Number* contrast, indicating that the difference in ratings between the *Ordinal* and *Different* conditions was significantly smaller in the *Number-Duration* than in the *Size-Number* condition.

Table 3.7. Aggregate analysis of Experiment 10. Significance tests for batches of coefficients in the ANOVA summary table and the individual *t* coefficients given by the Satterthwaite approximation. Abbreviation of interaction coefficients follows from first letters of the main effect terms: Ordinal vs. Different becomes OvD and Number-Duration vs. Size-Duration becomes NDvSD.

ANOVA summary	Coefficient	*B*	*SE*	*df**	*t*	*p*
	(Intercept)	-0.01	0.07	64.67	-0.12	0.9085
F(2,72.07)	Ordinal vs. Different	-1.03	0.15	68.27	-6.67	0.0000
=39.34	Ordinal vs. Ratio	0.99	0.12	69.25	8.46	0.0000
F(2,64.62)	Number-Duration vs. Size-Duration	0.45	0.17	70.04	2.69	0.0089
=3.68	Number-Duration vs. Size-Number	0.13	0.16	61.54	0.80	0.4271
F(4,68.98)	OvD:NDvSD	-0.21	0.38	75.13	-0.57	0.5722
=3.029	OvR:NDvSD	0.31	0.29	81.66	1.07	0.2890
	OvD:NDvSN	-0.88	0.36	64.32	-2.44	0.0173
	OvR:NDvSN	0.02	0.27	62.31	0.08	0.9338

3.20 Interim discussion

Experiment 10 showed that, like Experiment 9, subjects rated the *Ratio*, *Ordinal*, and *Different* sequence types differently. Most importantly, subjects rated the *Ratio* sequence pairs highest, indicating that they found them to be most similar. Thus, the ability to extract ratio representations of sequences and compare them across dimensions extends to size, duration and number.

The difference in performance between the number-duration condition and the others is difficult to explain. If the effect is not due to sampling error, it could be that the fewer numbers of trials in that condition led to a lack of experience in the task and thus difficulty differentiating sequences that were clearly different from each other, but this should have affected performance for the *Ratio* sequence types in that condition when it did not.

3.21 General discussion

3.21.1 Summary of results and general conclusion

To summarize our results, we have demonstrated that subjects can compare sequences on the basis of the amount of information preserved about relative magnitudes when transitioning from the first sequence to the second. This was true in the case of within-dimension comparisons of noise brightness, noise loudness, object height, and object size and for cross-modal comparisons between noise loudness and object size and between noise brightness and object height. This was also true when comparing sequences across the canonical magnitudes of space (object size), time (multimodal interval duration), and number (Arabic numerals).

Our results indicate that adults automatically extract relations based on inter-item ratio and inter-item rank information and use these abstract patterns to compare sequence patterns within and across modalities and across different dimensions of magnitude within-modality. Moreover, patterns that preserved inter-item ratio information were rated as more similar than patterns that only preserved inter-item rank information. This is consistent with our prediction that a dimensionless representation of a ratio scale supports the ability to reason about abstract magnitudes within and between dimensions.

Further, consistent with our speculation that the system of musical contour representations may be connected with generalized representations of magnitude (Bonn & Cantlon, 2012), we have demonstrated that similar behavioral signatures underlie comparisons of sequences in the auditory domain, visual domain, across modalities, and across magnitude dimensions.

One concern about the spontaneity of ratio representations arises from the fact that the tasks in this set of experiments require an explicit response about their similarity, which may be the trigger for generating the sophisticated analogical reasoning behavior. Though there is certainly evidence for subjects' adaptation to the task over time (see the difference between the two, separate, within-dimension comparisons for noise brightness), an exploratory analysis of first-trial behavior pooled across the first experiment revealed the same ratings pattern. In addition, there is no a priori reason why similarity ratings would trigger sensitivity to the most important difference demonstrated: that between *Ratio* and *Ordinal* sequences.

Future work will need to explore the extent to which fine-grained representations of relative magnitude underlie performance in dual-magnitude-judgment tasks as well as in tasks that require binding together stimulus values in memory for reproduction or categorization. It may be that unique, additional mechanisms idiosyncratic to pairs of dimensions contribute to varying levels of success in those tasks and future studies will need to tease apart contributions from both sources. For instance, for humans, event duration may be a less salient cue in general when paired with size cues, leading to asymmetrical interference between dimensions.

Additional work is needed to determine whether ratio representations are generated by calculations specific to each magnitude dimension independently and subsequently compared (implicit representation), or whether ratios are calculated by a centralized module that receives converging inputs from individual magnitudes (explicit representation), and whether these calculations take place at early or late processing stages. However, for these pursuits to be fruitful, the level of analysis at which these research questions are posed as well as the hypotheses themselves will have to be clarified considerably: ‘explicit’ and ‘implicit’ have different connotations when applied to the computational level of analysis (what problem is the system solving?) as opposed to the algorithmic and implementational levels of analysis (how the system is solving them; Marr, 1982). A centralized representation of a latent variable in a model framed at the computational level—for example, in an acyclic, directed graphical model—may be calculated in a distributed or implicit fashion at the algorithmic or implementational levels.

We focus our brief, speculative discussion of this question at the computational level. For example, one speculation might be that a shared or centralized representation of ratios is more appropriate when individual sensory/perceptual representations of magnitudes in each domain arise from common environmental causes, leading to correlated ratios among sensory inputs and a justification for efficiently representing a joint distribution over ratios with a common factor (a form of dimensionality reduction or the so-called blessing of abstraction). Separate representations of ratios might be more appropriate when the observer needs to keep causally unrelated representations independent: causally unrelated variations in ratios may be similar but are not a generated by a common source and thus there is no rational justification for combining information from multiple dimensions. In the next section, we discuss different causal scenarios that might generate mappings between ratios and scenarios that call for potentially different types of calculation.

3.21.2 Why Ratios?

Why would the nervous system need to spontaneously represent ratios *and* subsequently generate a mapping across ratios generated by different magnitude dimensions? While Vallentin et al. (2012) speculate on a handful of reasons why within-dimension ratio representation may be important for survival, the question of cross-dimension ratio mapping is not easy to answer, save for a few special situations. Different kinds of causal structures in the environment generate proportional changes in magnitude across dimensions. For example, one possibility is that rate processes generate quantities that are proportionally related. The simplest example is the equation Distance = Rate × Time, in which the spatial quantity of distance is proportional to total duration, given a constant rate. However, this example does not extend to all types of mappings of proportions across magnitudes. For example, larger animals tend to make louder sounds (object size to noise loudness mapping), but this does not have anything to do with an underlying rate of motion. Thus, there may be more than one way of generating mappings of ratios across domains and each one may require different representations.

If there is no single environmental scenario that generates cross-dimension ratio mappings, then why do subjects succeed in all the tasks reported in this chapter? The generation of ratio and relative-magnitude representations and spontaneous mapping of those ratio representations across dimensions in our experiments may have more to do with sophisticated analogical mapping than it does with the mind's way of representing similar causal structures in the environment. That is, mappings or isomorphisms between representations need not have *direct* roots in problem solving (either through evolution or ontogenetic processes) of specific computational problems that the environment presents.

The immediate benefits of spontaneously generating mappings between dimensions remain elusive. One possibility is that generating mappings in other dimensions facilitates or expands reasoning capabilities available in one domain to others and that this provides an overall problem-solving advantage to a given individual. However, this solution is not specific to reasoning about magnitudes; for example, mathematical proofs of propositions in set-theoretic language can sometimes be more easily understood and proven when mapped to a geometric domain (see Nagel & Newman, 2001, for dramatic examples). Further work will be necessary to determine if the ability to generate mappings across ratios depends on overall intelligence and memory capacity.

Chapter 4. Transfer of distributional learning across magnitudes

4.1 Preview

A generalized system of magnitude representations may potentially be characterized as a prior expectation that magnitudes in the environment are correlated, particularly when it is also assumed that those magnitude values have a common environmental cause. Thus, parallel to behavior observed in multisensory perception, observers equipped with such a prior should be able to use the statistical properties observed in one magnitude to predict magnitude values in unobserved, but causally related magnitudes. Previous studies have confirmed this intuition at a very coarse-grained level (i.e., transfer of ordinal representations via analogical reasoning), but we show that subjects can transfer detailed knowledge of stimulus distributions across magnitude domains in a supervised statistical learning experiment.

4.2 Introduction

What computational problem is the mind solving when it generates abstract magnitude representations? Walsh (2003, see also Bueti & Walsh, 2009) conjectured that magnitude representations need to be combined in preparation for action, but this is too vague a conceptualization to be theoretically helpful or to make any precise, computational predictions. Other authors have suggested that different kinds of magnitude might be correlated in the natural world (de Hevia, Izard, Coubart, Spelke, & Streri, 2014), which would lead to some kind of structured prior representation of all the pairwise correlations across different kinds of sensory continua. This is similar in spirit to the concept of amodal representations in multisensory perception, in which some modality-independent property of the world generates correlated cues across the senses. This conceptualization is currently too imprecise when characterizing abstract magnitude representations because the causal source of potential correlations is left unspecified—perhaps because it is too difficult to specify them in any complete or concise way to gain a sense of their generality.

However, a prior over correlations among magnitudes makes intuitive sense because some kinds of quantity are necessarily related. For instance, given a constant rate of motion, the distance traveled on a path will be proportional to the time it takes to complete it. A totally different scenario involves the relationship between size and loudness: a wolf and a grasshopper mouse both produce howling sounds, but the wolf's howl is louder and lower in pitch than that of the grasshopper mouse, whose howl must be amplified and lowered in pitch to be audible to humans.

The intuitive conclusion one might draw from these scenarios is that sensory measurements on one magnitude dimension would allow prediction or imputation of unobserved values on another. This characteristic is similar to that found in studies of multisensory or amodal representations, where learning in one sensory modality seems to transfer to another via some abstract representation, such as a set of binary features (eg., Yildirim & Jacobs, 2013).

Previous work has demonstrated that infant subjects can transfer coarse-grained knowledge about abstract magnitudes from one domain to another. For example, Lourenco & Longo (2010) showed that infants can transfer something like a more-vs.-less representation of amounts among the domains of length, numerosity, and duration. In another study, de Hevia & Spelke (2010) similarly showed that infants transfer representations of increasing or decreasing values from numerosity to length.

Another interpretation of this kind of magnitude transfer is that it is generated by a type of coarse-grained analogical reasoning about ordinal relations among magnitude measurements (eg., Gentner & Medina, 1998; Lu, Chen & Holyoak, 2012; Chen, Lu, & Holyoak, 2014). In this case, inter-dimensional mapping need not be associated with any causal relationship in the environment or any resulting correlations. It could simply be that magnitudes share a common, abstract representation that facilitates comparison or transfer of relative magnitudes across domains.

However, it is unknown how detailed these abstract representations transferred from one magnitude domain to the other are. In a web-based experiment in adults, we tested how well subjects transferred distributional learning of categories over a magnitude continuum to a novel continuum at test.

4.3 Experiment 11: Transfer of distributional learning across modalities

To probe whether subjects can transfer fine-grained representations of magnitude statistics across sensory modalities, we devised a supervised statistical learning paradigm in which subjects learned to associate flashes of deep-sea anglerfish lights that varied in brightness and size with particular species labels. There were two different sets of distributions to which different groups of subjects were assigned. At test, subjects were given a sample of a sound that one of the anglerfish species would make and were then asked to produce samples of sounds they predicted could come from each species. The visual features and auditory features of the anglerfish were designed to be dynamically similar by fading in an out, thus suggesting a common causal source of the magnitude variation.

4.4 Method

4.4.1 Subjects

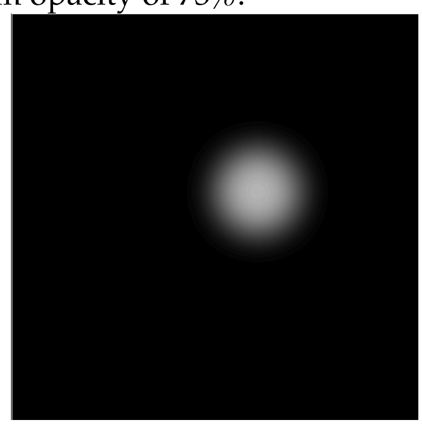
Subjects were recruited on Amazon Mechanical Turk (www.mturk.com) and restricted to have a 95% approval rating and IP address within the United States. We recruited 12 subjects for each condition; they were compensated $0.75 if they did not pass the sound check at the beginning of the experiment and $3.50 if they completed the full, 20-minute experiment. Recruitment continued until 24 total subjects had passed the sound check.

4.4.2 Stimuli

Visual stimuli were created in the web browser using standard HTML and javascript with jquery helper functions. Auditory stimuli were rendered in the browser via the WebAudio API, available in recent versions of Google Chrome, Opera, and Mozilla Firefox The viewing area was a 600-by-600 square with a 1-pixel, white border.

Anglerfish lights on each training trial were white circles of a particular diameter from the following sizes in pixels: { 50, 83, 117, 150, 183, 217, 250, 283, 317, 350}. Each size was matched respectively (from left to right) with a particular maximum opacity value: {0.20, 0.29, 0.38, 0.47, 0.56, 0.64, 0.73, 0.82, 0.91, 1}. Each fish stimulus edge was blurred for 10% of the fish's diameter using the CSS blending utility at the edges to simulate underwater distortion of light. Thus, the blurring effect extended fish lights 5% beyond the original diameter on each side. Each anglerfish light lasted for 1000 ms, with a 500-ms rise from 0% opacity to the pre-specified maximum opacity level and a 500-ms decay. See Figure 4.1 for an example fish stimulus.

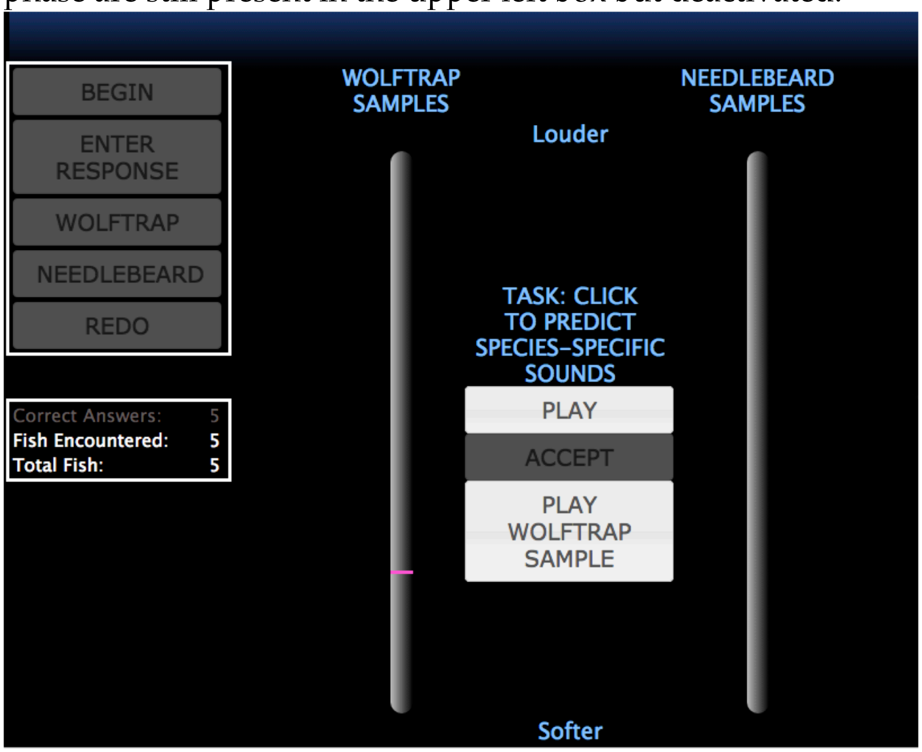
Figure 4.1. **Sample stimulus fish**. Diameter at 250px shown relative to 600-by-600-px border at maximum opacity of 73%.



Anglerfish sounds were generated by filtering a sample of white noise in the web browser; we manipulated the height of the bandpass filter and the overall gain of the sound based on subjects' input in the final phase of the experiment or from pre-selected values in the sound check. The sounds ramped to a maximum center frequency for the bandpass filter and a maximum gain with a 500-ms rise and 500-ms decay. The center frequency could range from 1000 to 10000 Hz and the gain control variable from 20% to 100% of the original volume.

Subjects controlled with a mouse the experiment flow by clicking buttons on the left-hand side of the screen. In addition, to help subjects keep track of their progress, trial tallies were displayed just above the button control panel. Subjects

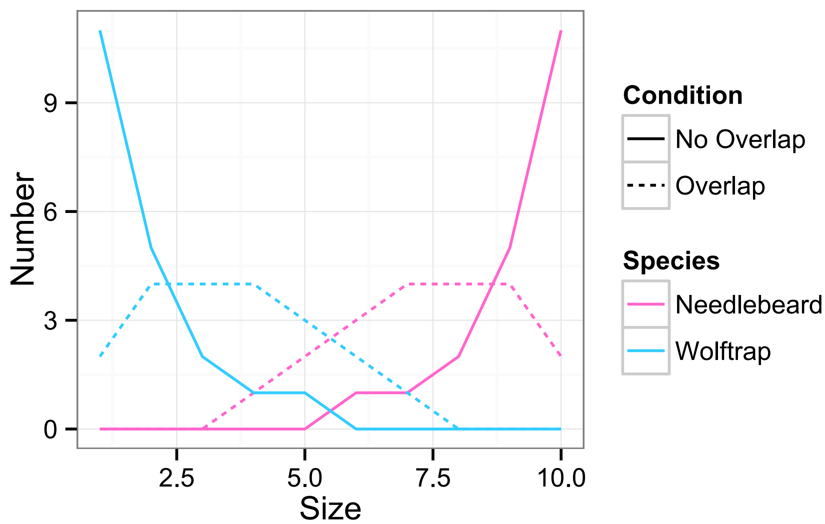
Figure 4.2. Sample generalization-phase display. Original buttons from the training phase are still present in the upper left box but deactivated.



could re-enter continuous size responses at their leisure if they made accidental clicks.

At test, subjects controlled their sound filter selections by clicking on one of two vertical continua (one for each species) and were forced to sample their selected sound value before moving on (see Figure 4.2 for a sample display). These were meant to be intuitive, visual representations of sound space. For each selection level, a horizontal, magenta bar appeared to mark the selection's place.

Figure 4.3. Stimulus distributions. Solid lines indicate 'Needlebeard' distributions, while dotted lines indicate 'Wolftrap' distributions. Pink distributions are from the 'Non-overlapping' condition while blue distributions are from the 'Overlapping' condition.



4.4.3 Stimulus Distributions

Stimuli in the training phase were drawn from one of two sets of two fish distributions, depending on the condition. The distributions of each condition varied in their means and variances and were created from idealized versions of probability density functions for beta distributions. In one condition, there was no overlap and a large separation between distributions ('Non-overlapping' condition) and in the other there was overlap ('Overlapping' condition). Distributions are pictured in Figure 4.3.

The numbers of stimuli present at each magnitude level are as follows for each distribution:

1. Non-overlapping, Small: {11, 5, 2, 1, 1, 0, 0, 0, 0, 0}
2. Non-overlapping, Large: {0, 0, 0, 0, 0, 1,1, 2, 5 11}
3. Overlapping, Small: {2, 4, 4, 4, 3, 2, 1, 0, 0, 0}
4. Overlapping, Large: {0, 0, 0, 1, 2, 3, 4, 4, 4, 2}

4.4.4 Procedure

The experiment consisted of 3 phases. The first phase was a sound check consisting of 10 trials. The sounds on each trial were filtered in the same manner as those in the test section, with rise and fall times of 500-ms each. Subjects heard two sounds and had to decide which one was louder, the first or second. To proceed to training, subjects had to choose the correct answer on 8 trials.

The second phase was the training phase. On each training trial, subjects clicked a button labeled 'Begin Trial' to initiate stimulus presentation. After an anglerfish stimulus was presented, subjects were then prompted to reproduce the size of the fish they just saw (to ensure they were paying attention). To reproduce the fish, subjects had to click in the viewing area and move their mouse. The distance between the current mouse position and the original click determined the radius and the brightness of the instantaneously rendered fish light. Then, they made a 2-alternative, forced-choice species judgment with one of two buttons labeled 'Wolftrap' or 'Needlebeard', the names of which are real species of anglerfish. Each choice was followed by feedback in text saying 'Correct!' or 'Incorrect!' and either a chime or buzzer sound, respectively. Each subject was exposed to 20, randomly ordered examples of each species.

During the test phase, subjects were instructed to produce 12 samples of guesses for sounds that each species might make based on one example of a wolftrap anglerfish (24 total guesses). Subjects could listen to this sample as many times as they wished during testing and could view the instructions whenever they needed a reminder. To do so, subjects clicked on one of two vertical continua that were 400-px tall (reminiscent of a slider from Chapter 2) to reserve a place for a single sample for each fish. They then pressed a button to hear their selection before either changing their selection or moving on to the next trial.

Audio samples were produced by calculating the relative position of the mark made on the slider and by calculating a proportionally equivalent point in log-frequency space and gain space between the minimum and maximum values indicated above (1000 Hz to 10000 Hz and 20% gain to 100% gain). Subjects were prevented from moving to the next trial until at least one attempt had been made to select a value. During this portion of the experiment, there was no feedback.

Subjects proceeded until the experiment ended automatically upon collection of 24 production samples.

4.4.5 Exclusion of subjects

We expected that subjects may perform at chance to game the system and decided to exclude anyone whose continuous size responses in training did not significantly correlate with the actual stimulus sizes.

Due to the open-ended nature of the testing phase, we expected to exclude people who did not separate category responses (i.e., seemingly responding at random or failing to use the stimulus space to get through the task without regard for instruction). Due to the small numbers of samples collected for each species for each subject, and because we were primarily interested in the differences between conditions rather than the general ability to transfer distributional knowledge across domains, we excluded people whose continuous responses were not marginally or significantly predicted by species label *in the correct, ordinal direction*.

4.5 Results

4.5.1 Training data

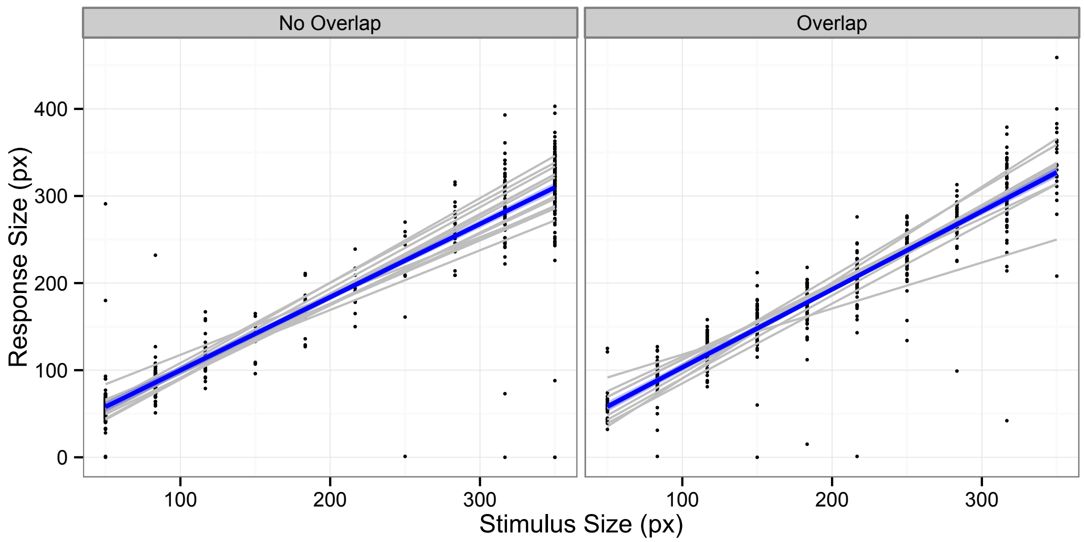
All subjects' size production responses were significantly correlated with stimulus sizes, precluding the need to exclude subjects on that basis. Data for continuous responses and categorization responses with regression lines as well as identification accuracy for visualization purposes are displayed in Figure 4.4.

One subject had to be excluded on the basis of giving one response in categorization for the entire training period. Figures 4.4a and 4.4b include all subjects, but for figure 4.4c, the subject who gave a single response is excluded.

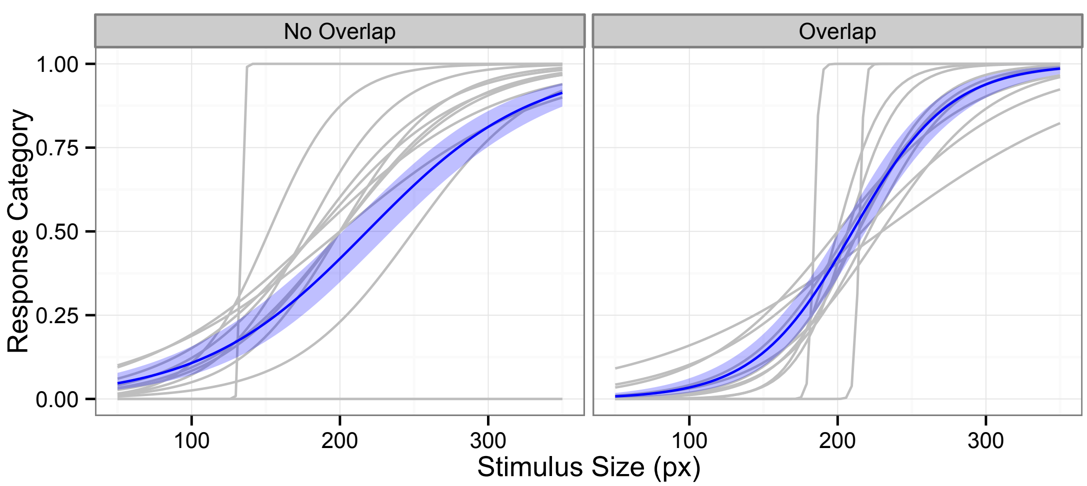
Figure 4.4. Training data. *a.* Individual data points represent one response, gray regression lines represent an individual subject's performance in the entire

training session. The blue regression line represents the overall regression line (for visualization purposes only; calculated with complete pooling of data). *b.* Categorization behavior as a function of stimulus size for all 40 training trials, with similarly plotted logistic regression lines for each subject and an overall regression line; category 0 is the smaller fish category. *c.* Average accuracy in categorization with bootstrapped 95% confidence intervals for the final 20 trials.

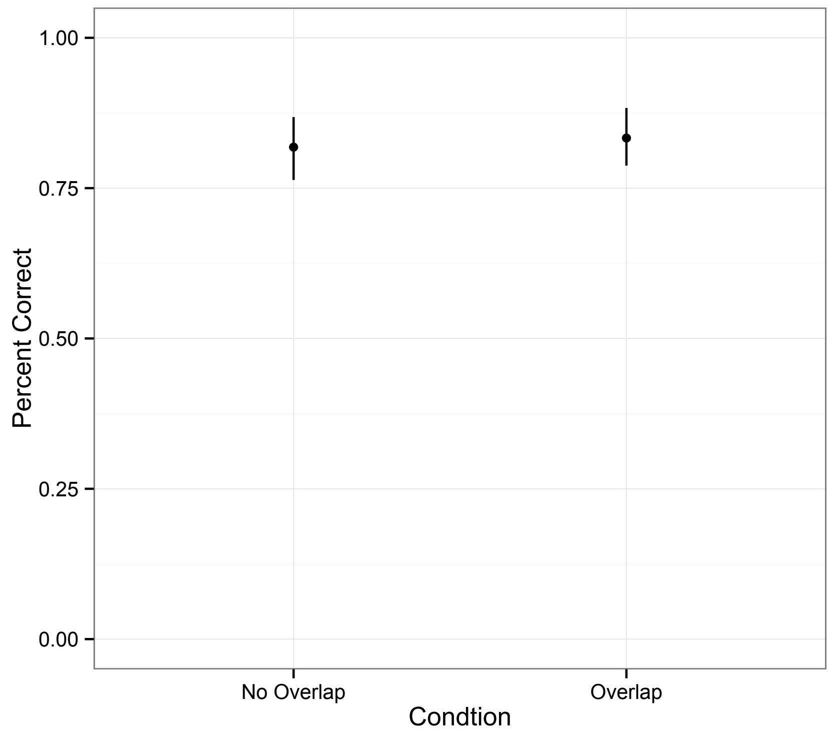
a.



b.



c.



Results show comparable performance between the groups. A multilevel, logistic regression with condition and a random intercept by subject as a predictor failed to show a difference in accuracy in classification of fish species (p = 0.99).

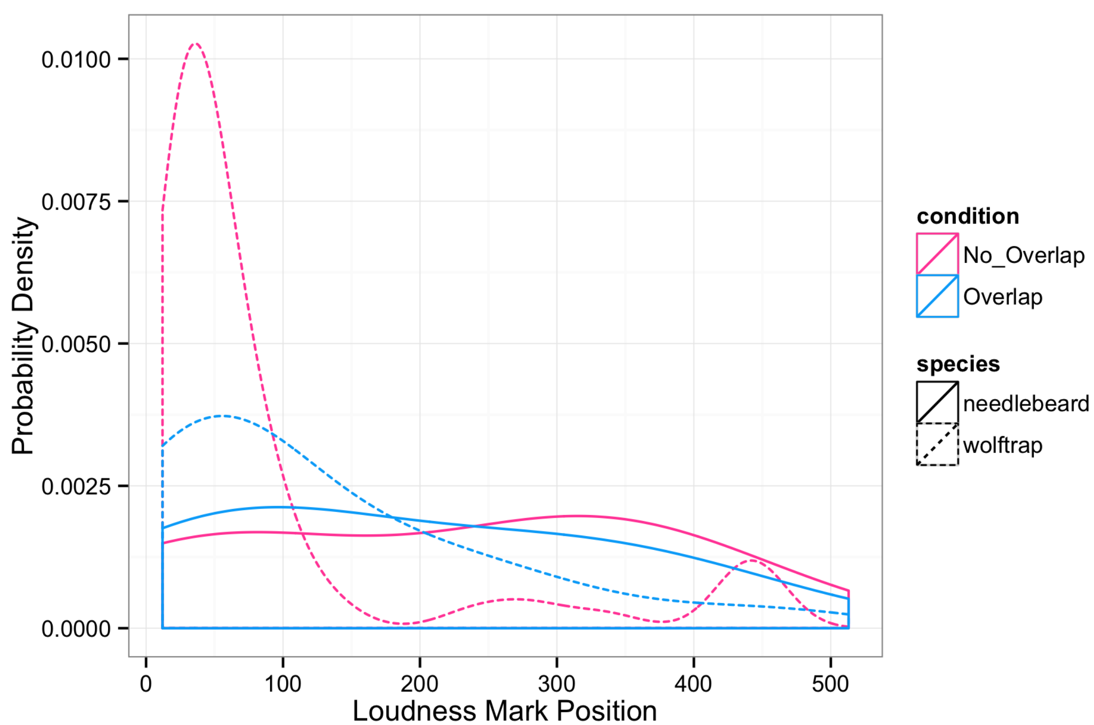
4.5.2 Test data

A summary of the qualitative results from all subjects is displayed in Figure 4.5. These figures display the qualitative differences in production performance in the auditory domain: those in the overlapping condition produced distributions of stimuli with larger amount of overlap between the two category distributions (Figure 4.5a) and distribution means that were closer together (Figure 4.5b). In addition, subjects had an overall, qualitative bias to choose quieter sound values.

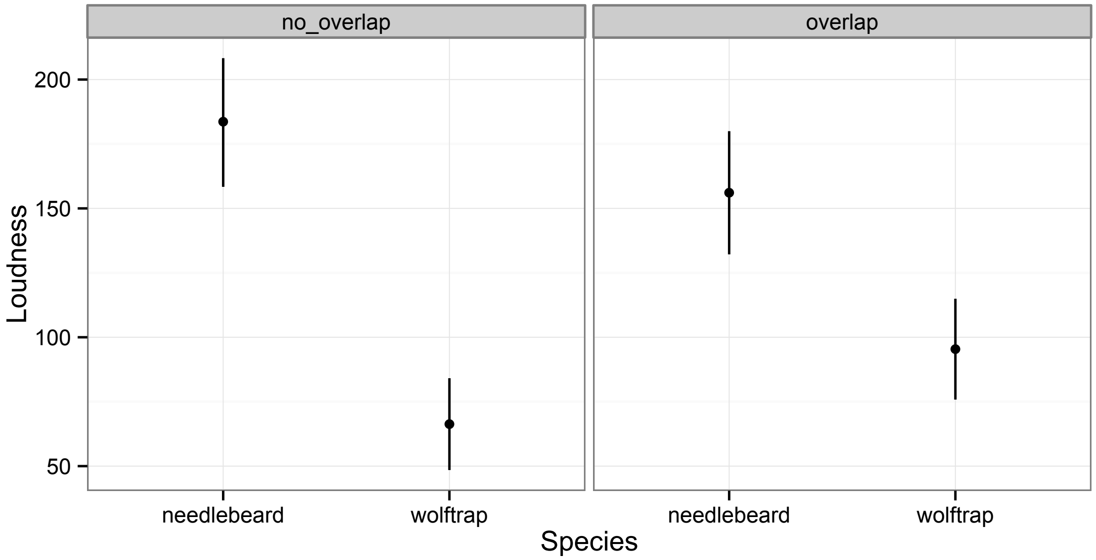
We excluded 4 people from the Non-overlapping condition and 6 people from the overlapping condition in the statistical analysis of test data according to the *a priori* exclusion criteria.

Figure 4.5. Generalization data. *a.* Raw, pooled distribution of responses displayed with a non-parametric density estimator, with different linetypes by species category and different line colors by condition. *b*. Bootstrapped 95% confidence intervals around category means for each species by condition, based on complete pooling of raw data values for visualization purposes.

a.



b.



A multilevel regression analysis with contrast-coded species and condition as predictors of raw pixel height values and Satterthwaite-approximated degrees of freedom revealed a significant effect of species (B = 178.27, SE = 22.64, df = 12.00, t = 7.87, $p < 0.0001$), no main effect of condition, indicating no difference in overall mean pixel location between conditions (B = -5.05, SE = 35.86, df = 12.00, t = -0.14, p = 0.89), and a significant interaction between species and condition (B = 111.87, SE = 45.29, df = 12.00, t = 2.47, p = 0.0295).

The predicted means of the regression models also revealed that the ratio between the distances between means in the responses across the non-overlap to overlap conditions (1.918) was very close to the empirical ratio of the distances between category means in training (1.947).

4.6 Discussion

The qualitative and quantitative results revealed that subjects transfer category-distributional knowledge between magnitudes from the visual domain to the auditory domain, given the highly constrained design of the experiment. Moreover, the ratio between the category-mean differences for each condition at test closely matched the ratio between differences in category means in training. The results suggest that abstract magnitude knowledge is more detailed than a simple transfer of ordinal knowledge of category means across modalities.

However, though we forced subjects to hear the stimuli they chose in the test phase, it is possible that subjects performed the task entirely visually, choosing samples based on relative location on the slider. Even if this were the case, this would indicate a transfer of magnitude knowledge to some abstract domain in which relative location in psychological space is visualized. Further work will be needed to tease apart these possibilities, including the development of a task that does not require the placement of continuous response values in a spatial mapping.

In addition, further work will be necessary to determine whether the spontaneous ratio-scale abstraction demonstrated in the previous chapter are the same representations responsible for the transfer of distributional knowledge in this task and whether the causal connection between the two tested stimulus domains is necessary for successful transfer. It could be the case that production responses are generated entirely by a sophisticated form of analogical reasoning (as in Chapter 3), rather than by an expectation for causally related dimensions to be correlated.

If the representation used here were generated by an expected correlation, other types of behavior associated with these multisensory representations should be confirmed with magnitude representations. Transfer of information across modalities is characteristic of multisensory or amodal representations, but is only one of several well-documented characteristics. For example, there are statistical advantages to combining cues from multiple domains in estimating latent properties of the world—i.e., to reduce noise in estimates (see Yildirim & Jacobs, 2012, for a review). In addition, learning or training in multisensory settings improves performance on unisensory tasks to a greater extent than training in the unisensory tasks themselves (Seitz, Kim, & Shams, 2006; Shams, Wozny, Kim, & Seitz, 2011). It is an open question as to whether combining cues from multiple magnitude domains would work in this same way, as the latent properties of the estimated, abstract, latent variables to be estimated as well as their relationship to the individual magnitude measurements is not fully understood.

For example, in multisensory (or even unisensory, multi-cue) integration, the causal or network structure of the model is clear: some latent property of the world generates the sensory measurements, and thus the sensory measurements are conditionally independent. (That is, given the latent property of the world, one sensory measurement provides no information about the other sensory measurement.) The causal structure in the present experiment is ambiguous. In our experiment, we trained subjects on arbitrary category boundaries corresponding to toy distributions of values along a magnitude continuum, but it is unclear whether the transfer effect reflects learning of latent categories which produce conditionally independent observations in size, brightness, and loudness, or whether the sizes (and brightness values) of fish themselves are the sole source of variance in loudness.

Other kinds of causal structures generating correlated magnitudes may not have the simple network structure characteristic of multisensory representations. For instance, travel distances and travel times are related via movement rate, but an acyclic, directed graph structure cannot be applied here. Knowing something about any two values will allow an estimate of the third, which precludes the kind of conditional-independence structures successful in explaining behavior in multisensory learning studies. Similar problems are encountered in any situation involving rate processes, such as counting processes—which connect discrete events or objects to space, time, or both via density calculations. Thus, there is reason to be cautious about drawing any general conclusions about interactions between magnitude domains like space, time, and number being similar to those encountered in multisensory and unisensory/multi-cue inference and learning.

Chapter 5. On binding magnitudes together:
Not all representations are the same

5.1 Preview

The hypothesis that a centralized magnitude system controls spatial, temporal, numerical, and potentially other representations (Walsh, 2003) is consistent with evidence of interactions between pairs of domains, though it is unclear whether the computations generating those interactions are truly homogeneous. Here we examine the potential heterogeneity in the structure of the internal models that adults generate in a statistical learning task to probe the interaction between object size and duration as well as the interaction between number and horizontal location. These two types of interaction are among the most commonly cited pieces of evidence for a generalized magnitude system. We show that the computational solutions that subjects generate for the same statistical learning task differ between these two pairs of dimensions, suggesting idiosyncratic rather than generalized solutions to the magnitude-binding problem.

5.2 Introduction

The hypothesis that a generalized magnitude system (Walsh, 2003) governs interactions between spatial, temporal, and numerical representations of magnitude has remained as controversial as it is influential (van Opstal & Verguts, 2013; Viarouge & de Hevia, 2013). Primary sources of evidence for a common or shared magnitude representation come from behavioral and neuro-imaging studies that demonstrate Stroop-like interactions between simultaneously presented magnitude dimensions (see Cantlon, Platt, & Brannon, 2009, and Bonn & Cantlon, 2012, for a review).

Most of the current literature (outside the chapter on spontaneous ratio-representations in this dissertation) conflates as one idea at least four, distinct, possible definitions for a generalized magnitude representation. One emphasizes shared neural real estate: this is the idea that magnitude processing draws on "distributed and overlapping" circuitry in the brain—a common module or mechanism (Pinel, Piazza, Le Bihan, & Dehaene, 2004; Cappelletti, Gessaroli, Hithersay, Mitolo, Didino, Kanal, Cohen Kadosh, & Walsh, 2015). Another related idea, inspired by the fact that most magnitudes conform to Weber's law of ratio-dependent discrimination and Stevens' power law in magnitude estimation, is that magnitudes are processed with similar coding mechanisms, even if they rely on highly distributed neural circuits.

A third, which potentially overlaps with the previous two but is conceptually distinct, is the idea that the generalized magnitude system supports the binding together of representations in different dimensions of magnitude into a single, abstract representation that resembles amodal representations of multisensory stimuli (eg., intersensory redundancy hypothesis; Bahrick & Lickliter, 2002). Closely related is the idea that different dimensions of magnitude in the world may be positively correlated, giving rise to prior expectations of correlations among magnitudes across either ontogenetic or phylogenetic timescales (Bonn & Cantlon, 2012; de Hevia, Izard, Coubert, Streri, & Spelke, 2014). This prior expectation for causally related (and thus correlated) magnitudes could give rise to the tendency to bind together some pairs of dimensions but not others (eg., Srinivasan & Carey, 2010).

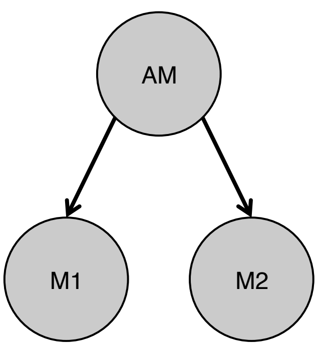
However, the literature on a mechanism for magnitude binding fails to provide any specific, computational models, though some authors have suggested that the generalized magnitude representation could resemble a system of cue combination (Lambrechts, Walsh, & Wassenhove, 2013), for which there exist many successful computational models that account for specific perceptual problems. Examples of these systems are frequently encountered in accounts of multisensory perception, the primary signature of which is that individual cues are weighted by their reliability (eg., Ernst & Banks, 2002; Battaglia, Jacobs, & Aslin, 2003; Körding & Wolpert, 2004; Alais & Burr, 2004; Bejjanki, Clayards, Knill, & Aslin, 2011). However, in these studies, the perceptual goal is usually taken for granted: 'reliability' is defined with respect to identifying a latent environmental cause, whether it be the location of an object in space or a native-language speech category. Common environmental causes provide correlated cues in multiple sensory systems, so observers should combine information from different modalities when it reflects a common environmental cause but segregate cues when it reflects different environmental causes (Körding, Beierholm, Ma, Quartz, Tenenbaum & Shams, 2007). It is not clear what the underlying cause(s) of correlated magnitudes might be in the natural environment or, more importantly, whether those causal structures, if they exist, warrant a generalized computational solution due to their similarity.

Instead of detailing the specifics of the computational problem that a generalized magnitude system might solve, debates in the literature have focused instead on arguing whether interference and bias effects in magnitude judgments arise at perceptual stages of processing or at later stages such as decision-making stages or during working-memory rehearsal. These debates have occurred in two separate domains. The first is in investigating interactions between time or event duration and space or object size (Yates, Loetscher, & Nicholls, 2012; Rammsayer & Verner, 2014; 2015). The second is in interactions between number and space or horizontal stimulus location, also known as the Spatial-Numerical Association of Response Codes or the SNARC effect, in which subjects display processing benefits when small numbers appear on the left and large numbers appear on the right (Dehaene, Bossini, & Giraux, 1993; Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006; Gevers, Santens, Dhooge, Chen, Van den Bossche, Fias, & Verguts, 2010). Evidence consistent with interactions between magnitudes occurring at later stages of processing is sometimes interpreted as evidence *against* a generalized magnitude system because task structures (and forced decisions or production judgments) vary from experiment to experiment, or at least provide a cautionary tale for making any strong claims about it (Opstal & Verguts, 2013).

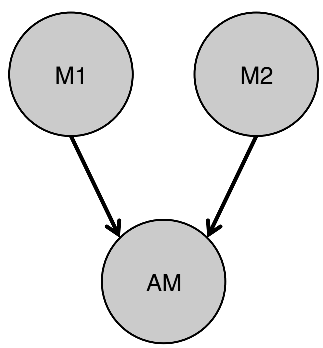
In this paper, we focus on what we view as a more fundamental question about the computational problem that subjects encounter in combining or binding magnitudes from different dimensions into a single, more abstract representation such as a categorical or ordinal representation. This binding process could occur at any level of processing from early perceptual processes to late-stage decisional processes such as handling potential actions or decisions in working memory. If a single, generalized magnitude system handles the process in the same way across pairs of magnitude dimensions, then we should observe the construction of similar computational solutions to the same learning problem presented with different pairs of magnitudes. On the other hand, the mind may handle each pair of magnitude dimensions idiosyncratically, ascribing different kinds of computational structures to essentially the same learning problem.

Figure 5.1. Potential graphical model types. 'AM' stands for abstract magnitude, while 'M1' and 'M2' stand for specific magnitude dimensions.

a.



b.

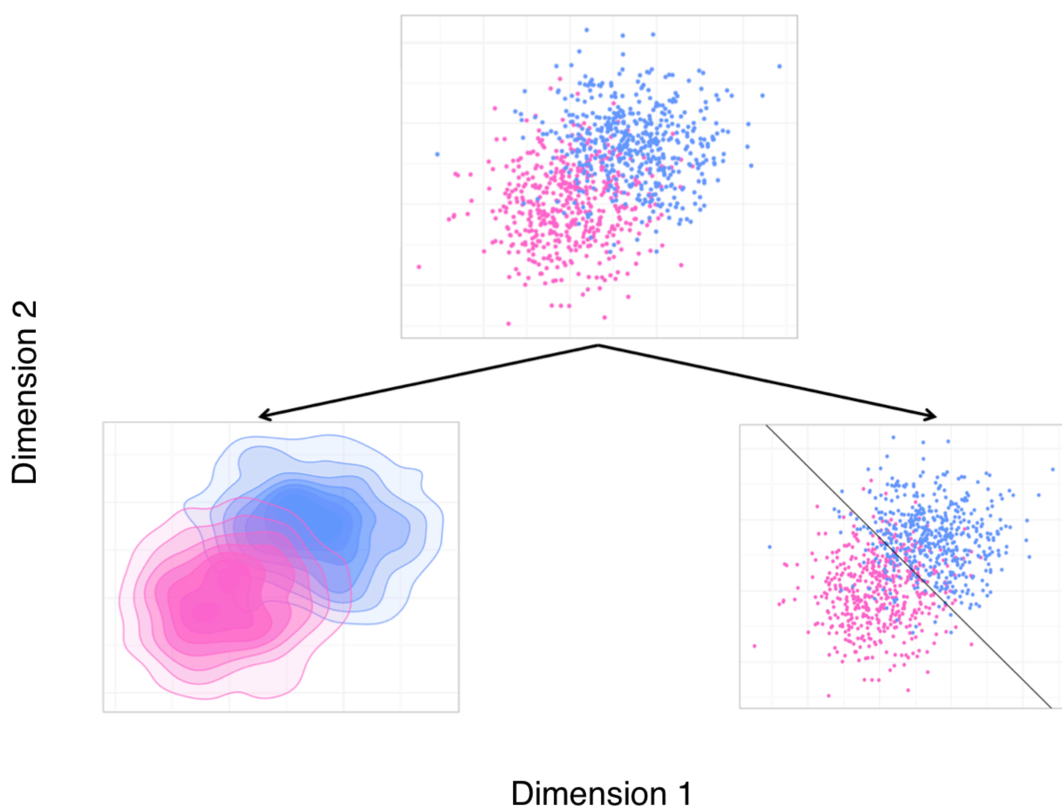


From a computational standpoint, if an abstract magnitude dimension plays a role in binding representations of pairs of dimensions together, the internal model could take one of two broad classes of form: the common-cause (also called common-factor) model or the common-effect model. In the common-cause model, sources of stimulation on two magnitude dimensions are viewed as arising from a single cause (see Figure 5.1a). In the common-effect model, values on two magnitude dimensions contribute or generate values on a common output dimension (see Figure 5.1b).

The common-cause model resembles latent-factor analysis models, whereas the common-effect model resembles multiple regression models (or, in the case of an unobserved output dimension, principal components analysis). Though the architectures appear very similar, they predict very different relationships between the two magnitude dimensions when the output dimension is observed. In the common-cause model, the individual magnitude dimensions are *conditionally independent*: that is, when the value of the causal variable is observed (or conditioning on the causal variable), the values on each individual magnitude dimension are independent. Observing one magnitude value, given the value of the cause, gives no further information about the other magnitude value. In the common-effect model, the individual magnitude dimensions are *conditionally dependent* or *marginally independent*, meaning that when the value on the abstract magnitude dimension is observed, it induces a dependency between the individual, contributing magnitudes. Note also that the common-effect model is of the same general structure as the feed-forward neural network models used to argue that numbers interact with space in working memory, which serves as a kind of 'output' dimension; in effect, these authors assume this model is appropriate for interactions observed in the SNARC effect. (For further references on conditional dependence and independence relationships expressed in Bayesian networks, see Bishop, 2006; Pearl, 1988; Pearl, 2000).

For an illustration of how these models might operate on data points from a 2-dimensional continuum, see Figure 5.2.

Figure 5.2. Two modeling strategies for correlated data points in the top panel. Colors indicate the underlying category. Left to right and bottom to top indicate increasing magnitude values on each of the 2 dimensions. In the bottom left panel, a common-cause model is illustrated: the idea is to model the source distributions in two-dimensional space directly; i.e., the data is modeled as having been generated by two normal distributions with means along the y = x line and uncorrelated variances within each distribution. In the bottom right panel, the common effect model is illustrated in one possible solution: drawing the best possible classification line (linear discriminant analysis). Other solutions like multiple logistic regression are also possible.



The question we asked was as follows: given exposure to two magnitude dimensions that map onto a third, categorical dimension, what is the nature of the mapping that subjects learn? Is it the same for distinct pairs of magnitudes? We investigated two of the most commonly used pairs of magnitudes whose interactions are cited as evidence for an abstract-magnitude system: (1) object size and stimulus duration and (2) number and space. We constructed a task in which encouraged subjects to learn that sample values on the two magnitude dimensions were correlated. Subjects had to make category judgments on exemplars falling within that 2-dimensional space and learn the locations of two, 2-dimensional distributions in that space.

5.3. Overview of Experiments

We devised a supervised, statistical learning task to probe the basic structure of subjects' computational solutions to a 2-dimensional categorization problem. Though we told each subject to attend to 2 dimensions to find a solution, subjects had to discover the underlying statistical structure of the categories and a decision-making procedure via trial and error.

We expected subjects to learn that the category means could be mapped to a single, abstract, magnitude dimension (eg., 'larger' or 'smaller') because the category means were placed along the $y = x$ line (see Figure 5.2c for a visualization) and the marginal distributions of stimulus values on each dimension were correlated. Within each category, the dimensions were uncorrelated. We increased subjective, structural ambiguity about the generative, statistical structure by placing a large amount of overlap between the two distributions belonging to the categories.

Though the true structure of the stimuli was a conditional-independence or common-cause structure, the subjects could solve the problem by viewing the dimensions as conditionally independent, given the category (abstract dimension or common cause model), or as marginally independent (conditionally dependent, given the category, or the common effect model).

To probe subjects' solutions after categorization training, we asked them to *produce* values from each category. Our predictions come directly from probability theory: If the two produced values for each dimension were correlated, after adjusting for the category, then it would indicate that subjects viewed the dimensions as marginally independent, suggesting a common-effect structure. However, if no residual correlation were present after adjusting for the category, then it would indicate that subjects viewed the dimensions as conditionally independent, reflecting a common-cause structure.

5.4 General method

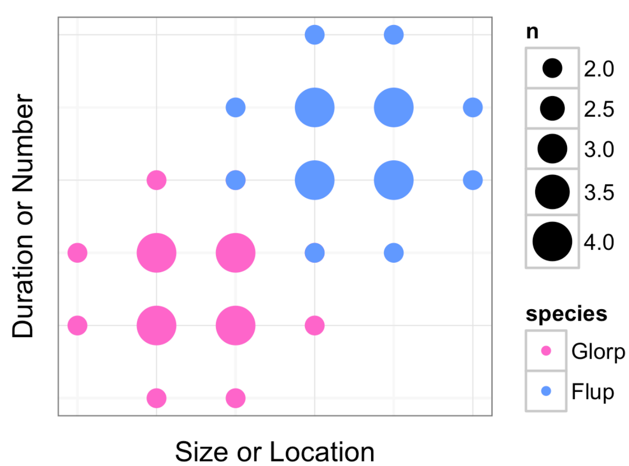
5.4.1 Experiment design

We trained subjects to identify two classes of alien species via their communication signals on a trial-and-error basis with stimuli that varied on two dimensions. The stimulus values along these dimensions for each species corresponded to two, overlapping, bivariate, unimodal distributions. The marginal distributions for each individual dimension (ignoring values on the other dimension and species label) and for the bivariate distribution (ignoring species label) were unimodal. The distributions were placed in the 2-dimensional stimulus space so that the stimulus values on each dimension were correlated. The dimensions *within* each species were not correlated. The distribution of stimulus values for each species given in each experiment is shown in Figure 5.3.

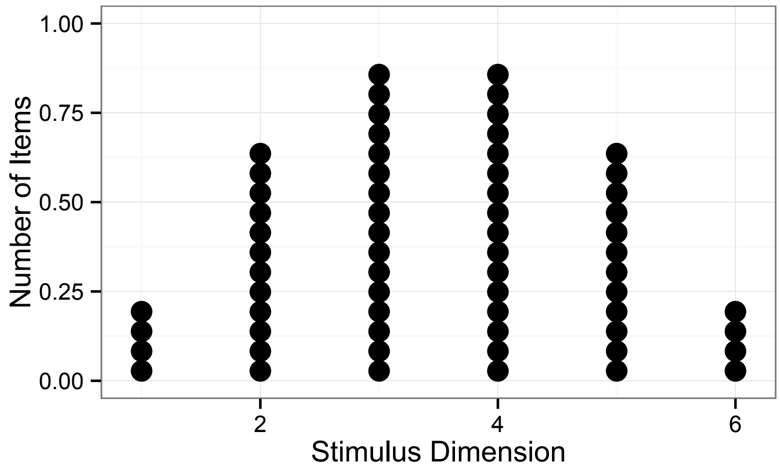
At test, subjects were given a species label on each trial and asked to produce a value on both dimensions in order to produce a signal communicate with that species.

Figure 5.3. Distributions of stimulus values. *a*. Increasing values on the axes correspond to increasing stimulus magnitude on an arbitrary log-linear scale. The dot size indicates the number of stimuli observed in that cell. Smaller dots = 2 presentations; larger dots = 4 presentations; color=species; 2 dots on y = -x line constitute overlapping species values. *b*. Marginal distribution of stimulus values along each dimension. Individual dots represent single stimuli. *c*. Marginal distribution of stimulus values (i.e. ignoring species label) in 2-dimensional space.

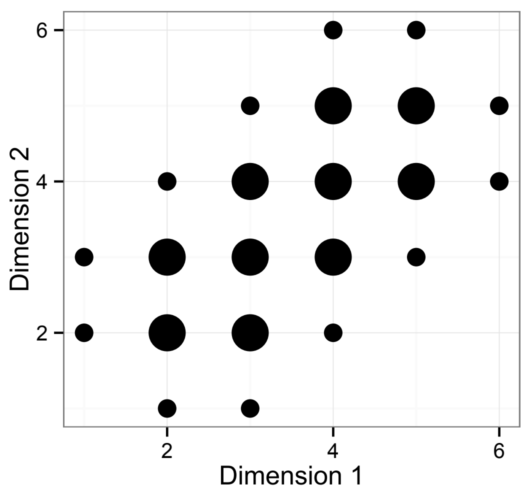
a.



b.



c.



5.4.2. Subjects

Subjects were recruited on Amazon Mechanical Turk (www.mturk.com) and restricted to have a 95% approval rating and IP address within the United States. We recruited 4 subjects at a time for each experiment until we reached at least 12 subjects passing inclusion criterion (see section 5.4.5 below) and each was compensated $5.50 for 30 minutes of work ($4 plus $1.50 bonus).

5.4.3 Stimuli

Stimuli were created in the web browser using HTML Canvas and the fabric.js library (www.fabricjs.com). The viewing area was an 800-by-600 black Canvas frame with a 1-pixel, white border.

Subjects controlled the flow of the experiment by clicking buttons on the left-hand side of the screen with a mouse or trackpad. In addition, to help subjects keep track of the flow of the experiment, trial tallies and point tallies (explained in section 5.4.4) were displayed just above the button control panel.

5.4.4 Procedure

On training trials, subjects clicked a button labeled 'Begin Trial' to initiate stimulus presentation. Then, they made a 2-alternative, forced-choice species judgment with one of two buttons labeled 'Glorp' or 'Flup'. Each choice was followed by feedback in text saying 'Correct!' or 'Incorrect!' and either a chime or buzzer sound, respectively; correct answers were rewarded with a single point and incorrect answers subtracted a point from the point total (only if the current number of points was greater than 0). Subjects were told that each point equaled an additional one-cent bonus to increase motivation for accuracy in a difficult task. Each subject was exposed to 32, randomly ordered examples of each species; thus, this block contained 64 trials.

On test trials, subjects were shown the name of a species and asked to make two continuous production judgments; specific methods for each experiment are discussed below. Subjects were prevented from moving to the next trial until at least one attempt had been made to select a value on each dimension. During this portion of the experiment, there was no feedback. To keep up motivation for maintaining accuracy, in the instructions subjects were told that they were silently gathering points for each signal correctly understood by the target species. Subjects were prompted with 25, randomly ordered labels for each species; thus, this block contained 50 trials.

At the end of the experiment, all subjects were rewarded with a $1.50 bonus, regardless of their performance.

5.4.5 Exclusion of subjects

Due to the difficulty of training and the open-endedness of the test phase, we included subjects in the analysis if they satisfied the following criteria. (1) At least one of the stimulus dimensions, had to significantly predict category choice in training. (2) The two dimensions of produced signals at test had to be significantly correlated, ignoring information about category. These two criteria indicated that the subject had performed well in learning the categories from stimulation and had successfully deduced the underlying correlation of stimulus values, irrespective of their relationship to the abstract category labels.

We anticipated that, in spite of the point system included to increase motivation in a difficult training task and an open-ended test, subjects would fail to pass both criteria at a high rate due to the open-endedness of the production task. In pilot experiments, some MTurk workers reported difficulty understanding directions, though most reported no issues. In addition, some

MTurk workers game the system by responding randomly in 2-alternative, forced-choice tasks (and even in magnitude production tasks) or seem to not care about performing well in learning tasks.

5.5 Experiment 12: Size and duration

In this experiment, we probed the relationship between representations of object size and duration of presentation. Each alien produced a signal that consisted of a solid circle that lasted for a particular interval.

5.6 Method

5.6.1 Stimuli

On each trial, subjects were presented with a light blue circle (hexadecimal color #33CCFF) in a random location within the HTML Canvas stimulus window. The circles could be randomly drawn from a uniform distribution over values in bins created from one of the following logarithmically spaced radius measurements (bin endpoints) in pixels: {[25, 35], [35, 48], [48, 66], [66, 91], [91, 127], [127, 175]}. The durations were drawn from a uniform distribution of values between the logarithmically spaced endpoints of the following bins in milliseconds: {[500, 692], [692, 956], [956, 1323], [1323, 1830], [1830, 2531], [2531, 3500]}. The endpoints were chosen to be proportionally equivalent in both dimensions, but absolute ranges of values were chosen to informally and roughly equate difficulty levels in estimation.

5.6.2 Procedure

In each trial in the first block, after clicking 'Begin Trial', following a brief interval of 900 to 1100 milliseconds, subjects viewed a stimulus in a random location in the viewing area. Response buttons became active after stimulus presentation. After the choice and feedback, the 'Begin Trial' button became active again.

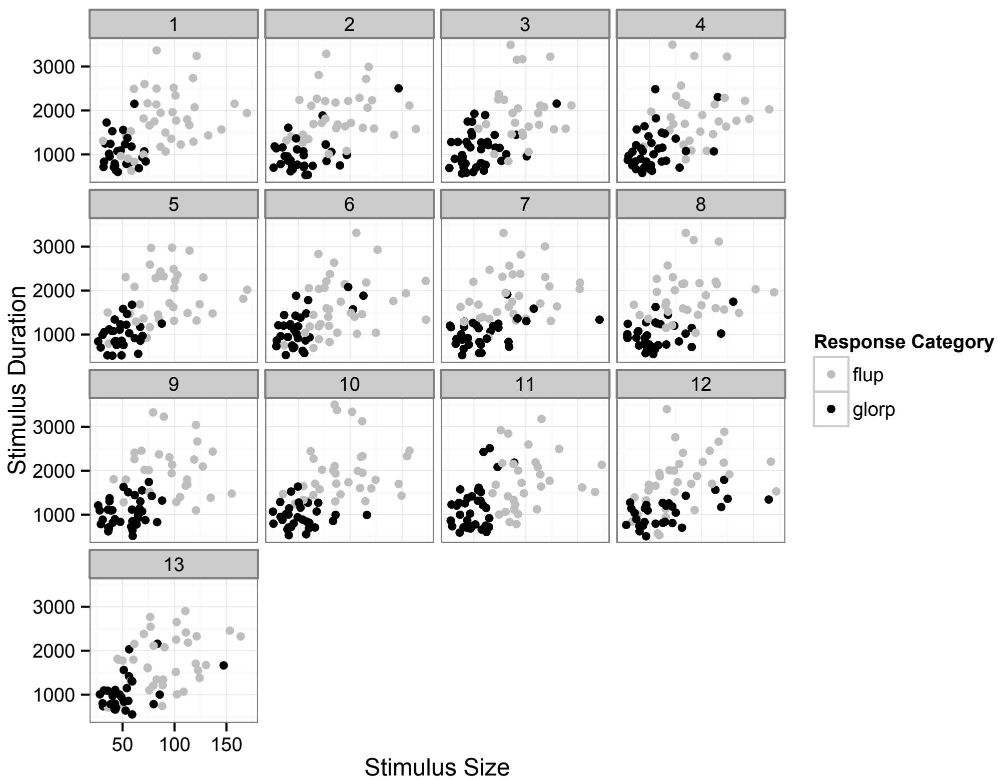
In each trial in the second block, subjects were presented with a species label at the top of the viewing area and a circle with a radius at the approximate, geometric mean of the trained stimulus values (42 pixels). To adjust its size, subjects clicked on a corner handle automatically generated by the fabric.js library. To record a duration value, subjects pressed a button to indicate they were ready to record, then clicked and held the mouse (or trackpad) for the desired duration. The object turned red for the duration of the click to add an additional cue to subjects that recording was in progress. For all experiments involving a duration production, subjects had the option of clicking a button to preview their stimulus. This was to remove some of the memory demand from duration judgments to make it more similar to the other production tasks in difficulty level. This could be done as many times per trial as the subject wished.

5.7 Results

Of the 20 subjects collected over the course of the experiment, 13 viable subjects were included in the final analysis in accordance with the exclusion criteria outlined in section 5.4.5.

Raw data from the training session are displayed in Figure 5.4. These data, which show a qualitatively successful separation between categories in the 2-dimensional space, demonstrate that these subjects successfully learned to classify the alien species.

Figure 5.4. Experiment 12 training data. Each subject's training data is presented in an individual panel, with stimulus radius in pixels on the x-axis, stimulus duration on the y-axis, and color reflecting the category choice for that stimulus.

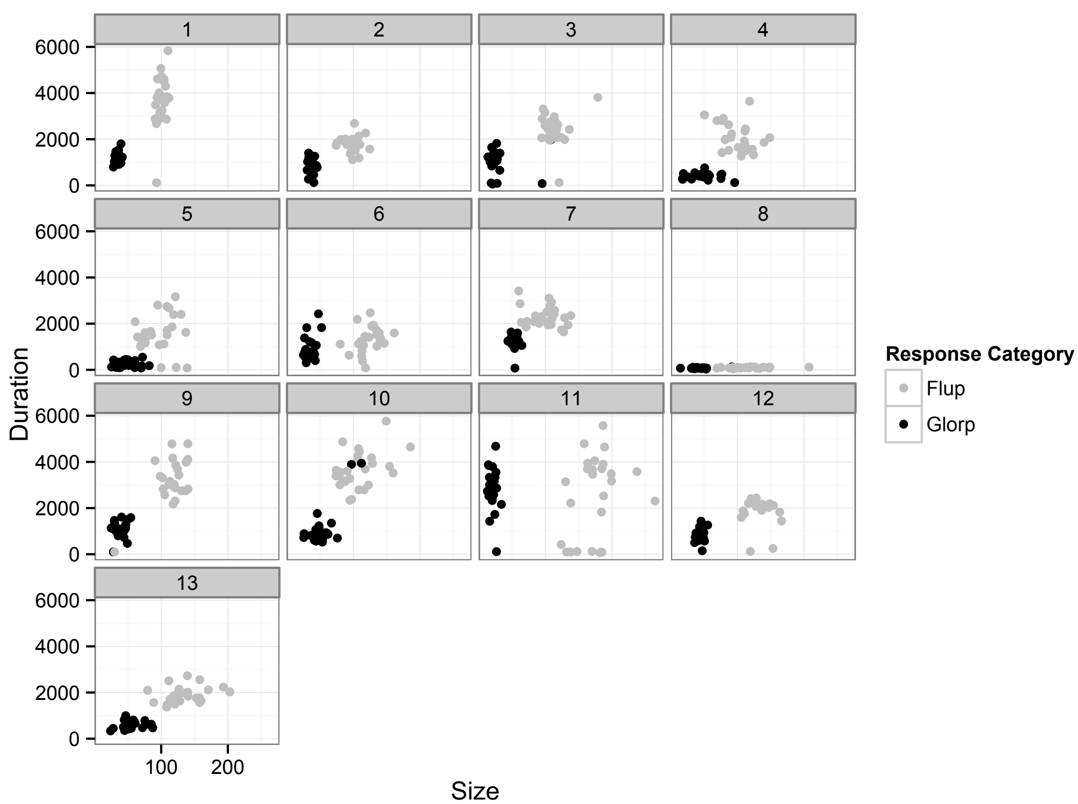


Raw data from the thirteen subjects on the production task is displayed in Figure 5.5. Overall, subjects displayed a high amount of variance in their use of the 2-dimensional response space. One subject's duration responses were overall shorter than other subjects' responses, though still correlated with object sizes; therefore these responses were still included in further analysis.

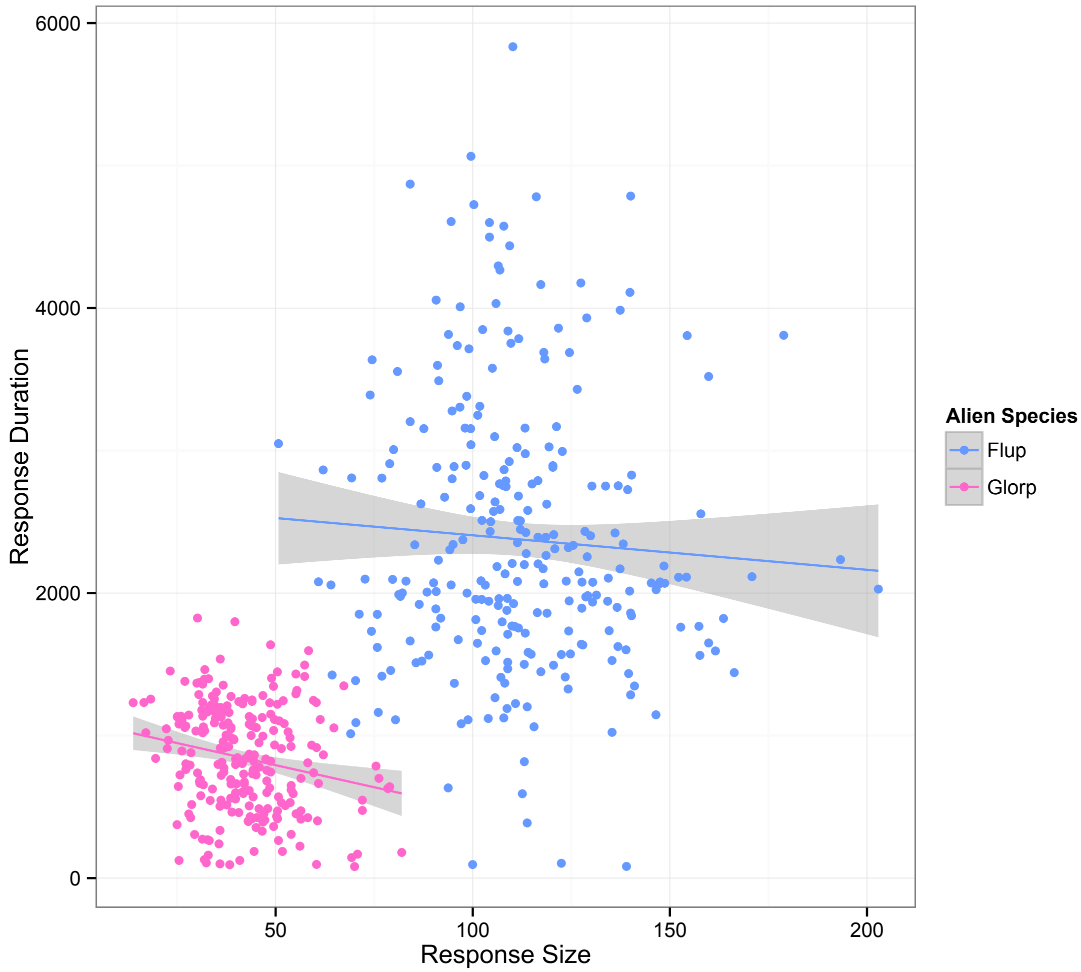
To analyze whether subjects responses for the two dimensions were correlated *controlling for/conditioning on* category, we performed a multilevel regression analysis predicting duration judgments in which contrast-coded species cue was entered as a predictor as well as object size. A random intercept term and a random slope term for size for each subject was included; no fixed interaction term for species and size predictor was included.

Figure 5.5. Experiment 12 production data. *a.* All data (including outliers excluded from the regression analysis) from included subjects is displayed, with produced size on the x-axis and produced duration on the y-axis. Color reflects the displayed category prompt on the given trial. *b.* For visualization purposes only: a scatterplot showing regression lines drawn over raw data (excluding two subjects with outlying ranges) with 95% confidence intervals; regressions reflect complete pooling of subject data.

a.



b.



To minimize the undue influence of outliers, we excluded all outlying points determined graphically in boxplots, whose default algorithm determines outliers by excluding points beyond 1.5 times the interquartile range. Outliers were particularly important to remove in this dataset because they could reflect incorrectly remembered category cues during production and thus produce spurious within-category correlations.

The analysis revealed the expected effect of species label (B = 1195.99, SE = 116.35, df = 576.3, t = 10.28, $p < 0.0001$; Satterthwaite correction for degrees of freedom). The analysis failed to reveal an effect of response size on duration (B = 2.07, SE = 3.22, df = 17.50, t = 0.65, p = 0.53), indicating subjects' within-category production values were not correlated.

5.8 Interim discussion

The results of this experiment show that magnitudes of size and duration are not correlated, controlling for category, suggesting that the relationship between size and duration in this experiment is one of conditional independence. This implies that to solve the task, subjects were using a common-cause model rather than a common-effect model: given the category, values on one dimension did not predict the other. This could also be understood in the following way: subjects likely learned that the dimensions were correlated at the level of the category, but not within category. Thus, once a sample has been generated by a particular species and that species is known, its value on size did not give any *additional* information about its duration.

5.9 Experiment 13: Number and space (line position)

In this experiment, we probed the relationship between representations of number and horizontal position on a line. This experiment was designed to test whether the structure of the internal mapping presumed to underlie SNARC effect is similar to or different from the internal mapping between size and duration. Each alien produced a signal that consisted of a 2-digit number at a particular location on the number line.

5.10 Method

5.10.1 Stimuli

On each trial, subjects were presented with a 2-digit number (hexadecimal color #33CCFF) in a random horizontal location above a 600-by-4-pixel, horizontal line of the same color centered within the HTML Canvas stimulus window. The positive integers were chosen randomly from uniform distributions over the following intervals ({[5,9], [10,14], [15,19], [20,24], [25,29], [30,34]}). The stimulus locations were randomly chosen from one of the following 6 sections of the number line, which extended from 100 to 700 pixels to the right of the viewing area boundary: {[100,199], [200,299], [300,399], [400,499], [500,599], [600,699]}. The exact pixel coordinate was chosen from a uniform distribution over pixel values within each bin. As in the numerical digit stimuli, this selection method was chosen to introduce task difficulty equivalent to stimuli chosen in the size or duration dimensions. Numbers were centered at the selected x coordinate.

5.10.2 Procedure

Categorization in the training block followed the same procedure as in the previous experiment.

In the training block, subjects selected numbers by entering a number in a text-input box and pressing ‘Record Number’. No default number was shown at the beginning of the trial, though a location placeholder was shown. This placeholder was a triangular, red tick mark just below the number line at the x-coordinate of 400 pixels to indicate the current location of the production stimulus to be created. Subjects could drag this tick mark along the length of the number line to indicate their desired location. When they selected a number, the number then appeared above the number line, centered above the upper point of the triangle.

5.11 Results

Of the 26 subjects collected over the course of the experiment, 13 viable subjects were included in the final analysis in accordance with the exclusion criteria outlined in section 5.4.5. Raw data from the training session are displayed in Figure 5.6.

These data show that, as with Experiment 12, these 13 subjects categorized aliens in the 2-dimensional space in the expected fashion.

Raw production data from the thirteen subjects is displayed in Figure 5.7. After inspection of this data, one additional subject (subject 9) had to be excluded from further analysis, as this subject reversed the locations of the categories at test, despite correctly identifying them in training. Outlying observations were removed using the same procedure used in Experiment 12.

Figure 5.6. Experiment 13 training data. Each subject's training data is presented in an individual panel, with stimulus number in pixels on the x-axis, stimulus location in pixels on the y-axis, and color reflecting the category choice for that stimulus.

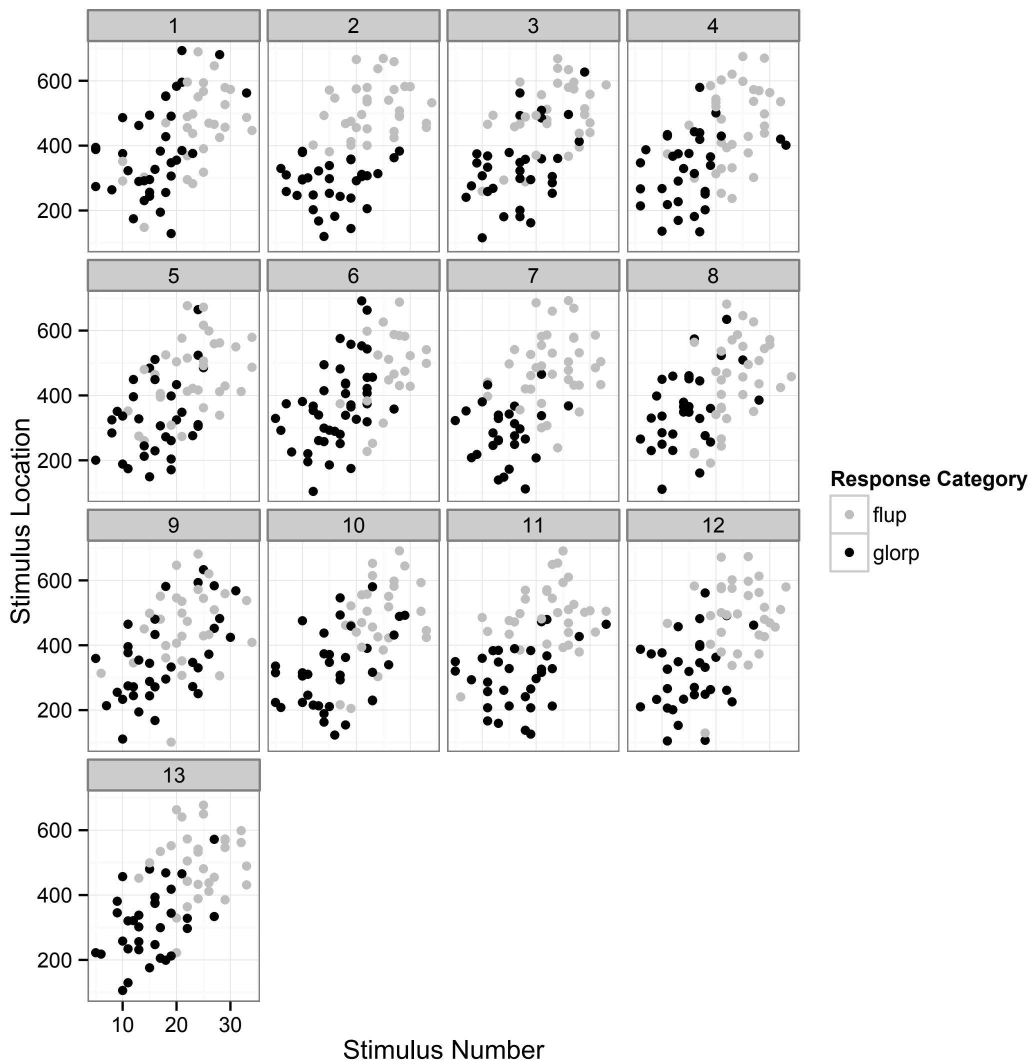
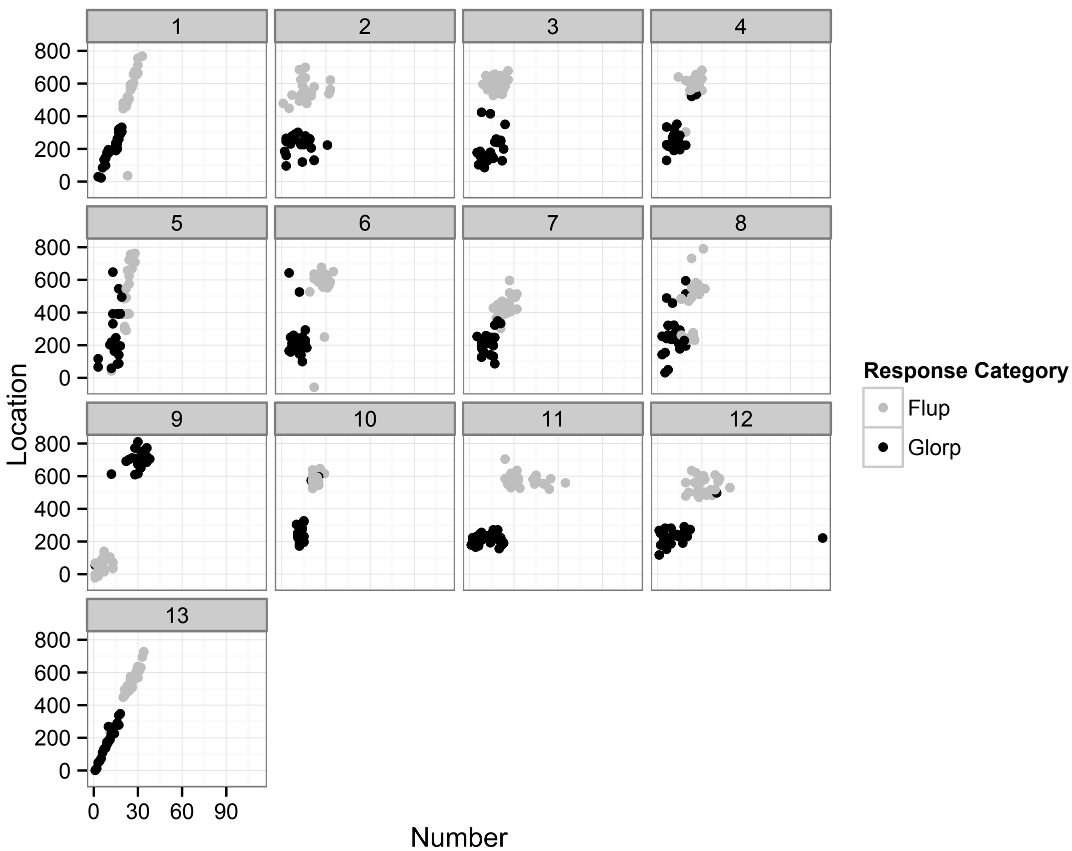
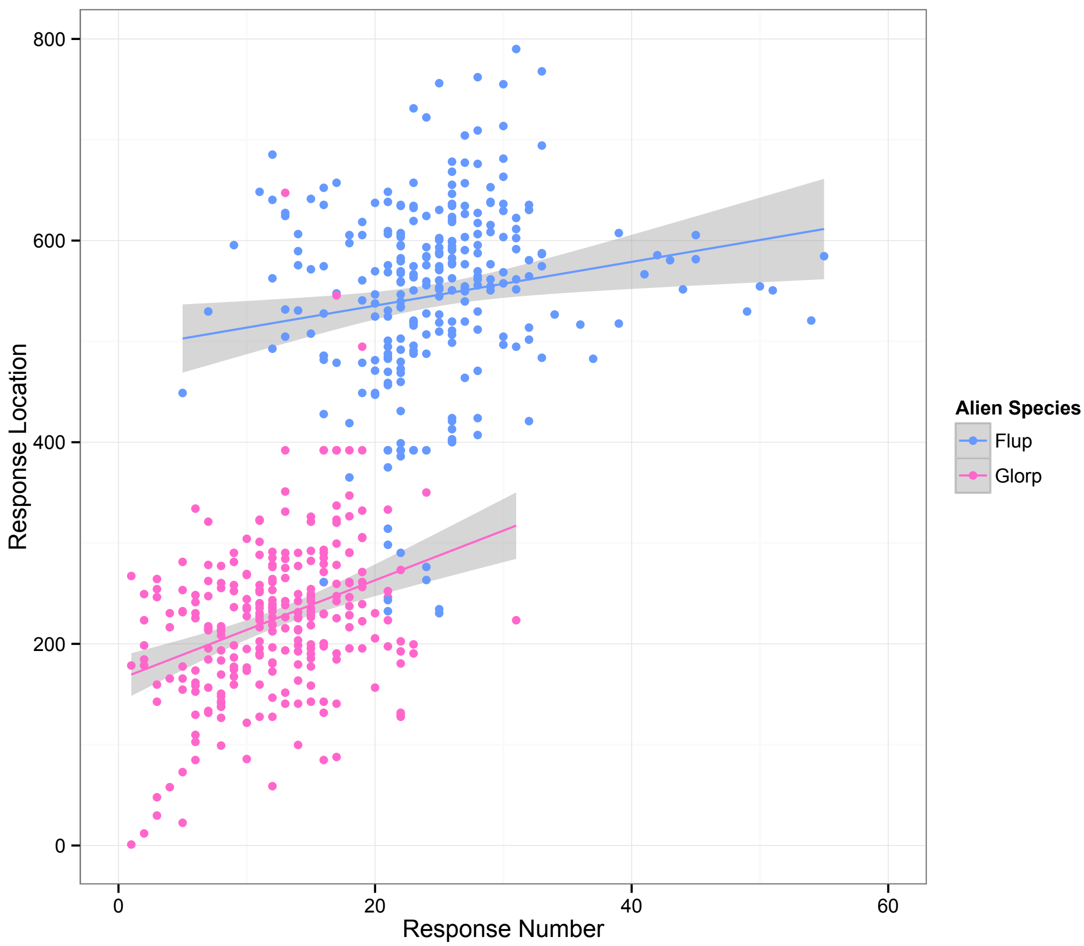


Figure 5.7. Experiment 13 production data. *a.* All data (including outliers excluded from the regression analysis) from 13 subjects passing initial exclusion criteria is displayed, with produced number on the x-axis and produced location on the y-axis. Color reflects the displayed category prompt on the given trial. Note the reversed category colors for subject 9. *b.* For visualization purposes only: a scatterplot showing regression lines drawn over raw data 95% confidence intervals; regressions reflect complete pooling of subject data.

a.



b.



Similar to the previous experiment, we conducted a multilevel regression analysis with contrast-coded species prompt as a predictor and response number as a predictor and response location as the dependent measure, with random intercepts and slopes for response number by subject. The analysis revealed the expected significant effect of species prompt ($B = 271.83$, $SE = 9.92$, $df = 533.70$, $t = 27.42$, $p < 0.0001$, Satterthwaite-corrected degrees of freedom). In addition, the analysis revealed an effect of response number on response location over and above category ($B = 3.97$, $SE = 1.39$, $df = 13.80$, $t = 2.85$, $p = 0.013$, Satterthwaite-corrected degrees of freedom).

5.12 Interim Discussion

The results for experiment 13 indicated that subjects' numerical responses were correlated with their location responses, even while conditioning on or controlling for the species label. This suggests that for number-location relations in this paradigm, subjects build (or modify a preexisting) common-effect model, despite the conditional-independence structure of the task and cover story.

Subjects 1 and 13 appeared to have qualitatively stronger results than the rest of the group, suggesting an alternative strategy like counting from left to right. However, no relationship between trial number and response location was found, indicating that these subjects may have imposed a stronger memory for the space-number mapping they used than other subjects. Moreover, the other subjects showed within-dimension correlation, albeit to a weaker degree, suggesting that even though there may be a categorical distinction in terms of spatial memory from trial to trial, they used similar number-location mappings.

5.13 General Discussion

To summarize the findings from the experiments, subjects' within-category responses were not correlated for the size-duration experiment but were correlated for the number-location experiment. This suggests that subjects construct a common-cause model in the size-duration mapping case but a common-effect model in the number-location mapping case.

The finding in the number-location case is consistent with the assumptions of neural network models of the SNARC effect, in which input dimensions of number and stimulus location are coded separately and feed into a decision-making variable.

However, the result in the size-duration case indicates something more like model structures that explain superior performance in multisensory learning paradigms (see Yildirim & Jacobs, 2012, for a review). In those models, latent features generate the observed sensory data. Further experiments exploring the relationship between size and duration should aim to (dis)confirm behavioral predictions of those models, including superior performance in unisensory tasks after multidimensional (rather than unidimensional) training.

Importantly, the experiments show that even under the same statistical learning paradigm, subjects learn different structures with different pairs of magnitudes, indicating that a single computational solution does not seem to be present for interactions between these dimensions. These are the pairs of dimensions most commonly cited as evidence for a generalized magnitude system, so these results suggest that the two lines of evidence may actually reflect different processing mechanisms, consistent with our post-hoc conclusions for Chapter 1. Moreover, arguments *against* a generalized magnitude system based on models of the SNARC effect (eg., van Opstal & Verguts, 2013) may not necessarily be validly used in evaluating interactions between other magnitude pairs. The size-duration pairing may be part of a generalized magnitude system, whereas the engine of the number-location mapping may have entirely different origins.

Why do the particular dimension pairs we used here elicit their respective models? The common-cause solution for the size-duration pairing could either have been predicted by a prior expectation of causal relationships between the two domains *or* that the subjects accurately learned the model they were presented with. The common-effect solution for the number-location pairing probably would have been predicted by any account that presupposed that numbers (digits) and locations on a line are independent; the mapping to the third dimension forces the dependence between dimensions which would otherwise be independent. In principle, if subjects learned the task well enough with sufficient exposure, they could discount their prior experience with number lines and accurately recover the conditional independence structure implemented in the current tasks. Conversely, if given a conditional dependence structure, subjects could potentially learn such a mapping in a size-duration pairing given enough exposure. In any of these cases, there is no reason to suppose that a *general* magnitude system in the form of a preset model type governs statistical learning.

Further studies will be necessary to carve out the space of models in different pairs of stimulus dimensions: how would number-duration mappings or number-size mappings behave with this learning paradigm? What would happen if dot arrays were used instead of Arabic numerals? What would happen with other dimensions such as brightness? More importantly, future work should examine in greater detail the causal structures in the natural world that generate the relationships between magnitudes and whether the types of models subjects create for different kinds of magnitude pairs in the lab reflect these environmental causal structures—and the extent to which these causal structures resemble one another.

One minor caveat for our results concerns further distinctions that could be made among models showing conditional-independence relationships, as in the Experiment 1. As Pearl (2000) explains, the common-cause model is indistinguishable from two other similar models when conditioning on the latent cause/abstract-magnitude variable: M1 → AM → M2 and M1 ← AM ← M2. These reflect a kind of causal path in which the abstract magnitude mediates the relationship between two individual magnitudes; for our purposes, we consider each of these alternatives as equally plausible for an abstract magnitude model.

In addition, we restricted our investigation to a situation in which subjects were forced to map two dimensions onto a third, categorical dimension. This reflects the most intuitive conception of 'generalized magnitude' in context of the idea that subjects may have a prior expectation that magnitudes should be correlated in the world: both the common-cause family and common-effect models should reflect a monotonic mapping onto a third dimension, which is the most commonly invoked conception of abstract or generalized magnitudes. There are other possibilities, as mentioned in Bonn & Cantlon (2012) and in the introduction of this dissertation, in which one magnitude dimension is mapped directly onto another via some (log-)linear or monotonic mapping, without an intervening or abstract variable except the mapping function itself. This would resemble a supervised, associative-learning model or univariate regression model. Function learning of this kind could be important for learning mappings between magnitudes in very specific kinds of environmental situations, but is unlikely to give rise to more general computational solutions across pairs of dimensions, given the results in Chapter 2.

Chapter 6: Concluding remarks

To summarize the findings of this dissertation, we have demonstrated the following in each chapter:

(1) A full matrix of comparisons of inter-dimensional bias effects in size, numerosity, duration, and brightness revealed no consistent patterns, suggesting that evidence for interactions between domains may not be the best evidence for a generalized magnitude system, at least in dual reproduction tasks. Two directions for future research were suggested: an exploration of the potential causal relationship underlying bias effects between each individual pair of magnitudes, and a comparison of the pattern of results with dual bisection (closely related to categorization as close 'large' or 'small') tasks.

(2) A set of studies examining comparisons of sequences of items varying in magnitude revealed that observers use abstract representations as fine-grained as ratio (as well as ordinal) relations in a domain- and magnitude-general way. This opens up the possibility that a generalized magnitude code may be about relative magnitudes rather than a common relationship among absolute magnitude values.

(3) Consistent with predictions of both the latent-representation account and the relative-magnitude code account of generalized magnitudes, learners can transfer knowledge of distributions over magnitudes belonging to categories of items from the visual domain (size, brightness) to the auditory domain (loudness).

(4) In post-tests after identically structured, supervised statistical learning tasks in two different pairs of magnitudes—size/duration and number/spatial location, subjects displayed patterns of behavior consistent with two different types of causal representation. Size/duration associations were represented in a manner consistent with 'common-cause' type models and number/space assocations were represented in a manner consistent with 'common-effect' type models. The results confirm the post-hoc speculation of chapter 1, but allow for an even stronger conclusion about the potentially idiosyncratic nature of relationships between magnitude pairs: even though the statistical structure of the samples subjects were exposed to was identical across pairs, learners imposed different model structures on their experience.

The take-home message of the experiments is that the most promising candidate for a generalized magnitude system among those formalized in this dissertation is one of dimensionless quantities derived from relative magnitudes (such as ratios and ranks). These quantities are at a level of abstraction that is sufficient for supporting comparisons of representations across different sensory and magnitude domains. In other words, ratios and ranks arise from comparisons among values within separate, absolute quantities, but qualify as abstract magnitude representations by themselves. In that sense, it could be difficult to distinguish relative magnitudes as an independent system of *representation* from one that is a system of possible arithmetic *operations* or a mechanism that performs comparisons on representations: they function as both because they are derived from operations over absolute magnitudes and appear to be a new, more abstract, kind of representation themselves. Further, operations could be performed on the representations derived from magnitude comparisons. This would predict that subjects could transfer compound representations of relative magnitude from one dimension to another—created by adding or subtracting ratios or ranks—and perform comparisons or equivalence judgments on those compound representations.

At a different level of analysis, it is unclear whether relative magnitude representations are merely separate codes that are generated by individual modules in the same output language or if it is a single, centralized, and perhaps even modular representation. Future work will need to explore these possibilities: for example, an experiment could use a method of selective adaptation to ratio values to explore whether adaptation to ratios transfers across domains; if that were demonstrated, it would be consistent with the possibility of a centralized representation that is shared across domains.

However, the lack of consistent patterns of results from chapters 2 and 5 does not rule out the possibility that some group of magnitudes may share a privileged relationship at a higher level of perceptual or conceptual analysis than required of these particular tasks or even as late as response generation. If that is the case, it would be imperative to examine the relationship between the system of relative magnitude representation and whether it underlies the computations necessary for generating responses in 2-alternative, forced choice tasks.

In addition, these results do not mean that space, time, and number do not share a privileged relationship in terms of computational resources or computational architecture. In the following section, we discuss additional possibilities for the ways these domains may relate.

6.1 Alternative types of interactions between space, time, and number

6.1.1 Inferences over rate processes

Walsh's original hypothesis concerned a shared representation or code for space, time, and number to explain interaction effects among these domains. This definition is sufficiently imprecise to allow for the possibility of a somewhat different interpretation: that space, time, and number share similar computational problems and thus recruit similar neural real estate. Not only might they share similar computational problems, but they may interact with each other in a specific way—in doing inference over *rate* or *intensity processes*, which are encountered in survival analysis, econometrics, and all branches of spatial statistics.

The most familiar versions of this problem concern interactions between space and time. For instance, the amount of time it takes to walk between two locations depends on the rate of motion of the walker. Inferences of this type may be over models of instantaneous rates of continuous motion or less sophisticated calculations of average rate (Average Rate = Distance / Time).

Less familiar than the relationship between distance and time are counting processes, which also concern a kind of rate calculation. An observer can calculate the average number of objects present across space, time, or both, which is a calculation of *density*. A more sophisticated inference, as in the continuous rate case, could be about the instantaneous probability of events or objects occurring at a given point in space, time, or space-time. The only situation in which density and instantaneous intensity are equivalent in stochastic processes (in contrast to constant-rate processes) is in the Poisson process, where events are uniformly distributed across space and/or time (Diggle, 2013). Many other examples of counting processes in which density and intensity are not equivalent occur: for example, when items repel each other in space (eg., trees in a forest or cell nuclei separated or repelled from each other by cell bodies) or when items cluster (foraging patches). Any animal that adapts to calculate these kinds of more complex counting processes stands at a predictive advantage. Moreover, some evidence suggests that the clustering of items affects numerosity estimates in humans (Ginsburg, 1991).

In particular, inferences over various types of counting processes occurring in space and time necessarily involve all three canonical magnitude domains of space, time, and number. Future work will need to address how much the shared real estate of space, time, and number is a reflection of the use of common encoding strategies as well as, or in contrast to, the need to integrate them for rate calculations.

6.1.2 Doing statistics over samples taken over space and time

Even more fundamental than rate calculations is the special kind of statistical inference required when including spatial or temporal variables as independent variables in a model. Events occurring in space or time tend to be autocorrelated: that is, the values of samples taken at different points in space or time are non-independent. Thus, statistical models of spatial, temporal, or spatio-temporal data must take into account with correction or explicitly model the causal source of interdependence among items in space (for a classic text on spatial statistics, see Cressie, 1993; for a classic text on time-series analysis, see Box & Jenkins, 2015). That causal source may be a type of rate or intensity process.

These considerations suggest that interactions between space, time, and number may reflect how closely they interact with computations of rate and intensity. Future work will need to expand on this possibility by exploring the manner and extent to which humans and other animals may be able to infer latent rates and instantaneous intensities.

6.2 Contributions of the present work

In conclusion, we have contributed detailed descriptions of two possibilities for a generalized magnitude system and provided evidence that relative-magnitude representations, rather than composite representations of absolute magnitudes, may be the best candidate for such a system. This work clarifies the space of hypotheses consistent and inconsistent with such a system and suggests several paths forward for designing studies that will further clarify the role of abstract magnitude representation in human—and perhaps non-human—cognition.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *1*, 257–262.

Austerweil, J. L., & Griffiths, T. L. (2011). A rational model of the effects of distributional information on feature learning. *Cognitive Psychology*, *63*, 173–209.

Austerweil, J.L., & Griffiths, T.L. (2013). A nonparametric Bayesian framework for constructing flexible feature representations. *Psychological Review*, *120*(4), 817-851.

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, *125*, 238–251.

Battaglia, P.W., Jacobs, R.A., Aslin, R.N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A*, *2*, 1391–1397.

Bejjanki, V.R., Clayards, M., Knill, D.C., & Aslin, R.N. (2011). Cue integration in categorical tasks: Insights from audio-visual speech perception. *PLoS ONE*, *6*(5), e19812.

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from arabic numerals: Developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, *74*, 286–308.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.

Bonn, C.D., & Cantlon, J.F. (2012). The origins and structure of quantitative concepts. *Cognitive Neuropsychology*, *29*(1-2), 149-173.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition*, *75*, 1–28.

Boroditsky, L. (2001). Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology*, *43*, 1–22.

Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*, 123–129.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C., & Ljung, G.M. (2015). *Time Series Analysis: Forecasting and Control* (5^{th} ed.). Hoboken, NJ: Wiley & Sons.

Brigner, W. L. (1986). Effect of perceived brightness on perceived time. *Perceptual & Motor Skills*, *63*, 427–430.

Brigner, W. L. (1988). Perceived duration as a function of pitch. *Perceptual & Motor Skills*, *67*, 301–302.

Brosch, M., Selezneva, E., Bucks, C., & Scheich, H. (2004). Macaque monkeys discriminate pitch relationships. *Cognition*, *91,* 259-272.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number, and other magnitudes. *Philosophical Transactions of the Royal Society B*, *364*, 1831-1840.

Butler, R. A. (1969). Monaural and binaural localization of noise bursts vertically in the medial sagittal plane. *Journal of Auditory Research*, *3*, 230–235.

Butler, R. A. (1971). The monaural localization of tonal stimuli. *Attention, Perception, & Psychophysics*, *9*, 99–101.

Cantlon, J.F. (2012). Math, monkeys, and the developing brain. *Proceedings of the National Academy of Sciences*, *109*, 10725-10732.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91.

Cappelletti, M., Freeman, E. D., & Butterworth, B. L. (2011). Time processing in dyscalculia. *Frontiers in Psychology*, *2*, 1–10.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*, 579–593.

Caviness, J. A. (1964). Visual and tactual perception of solid shape (Doctoral dissertation). Cornell University, Ithaca, NY.

Chang, A. Y.-C., Tzeng, O. J. L., Hung, D. L., & Wu, D. H. (2011). Big time is not always long: Numerical magnitude automatically affects time reproduction. *Psychological Science*, *22*, 1567– 1573.

Chen, J. Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, *104*, 427–436.

Chen, Q., & Verguts, T. (2010). Beyond the mental number line: A neural-network model of number-space interactions. *Cognitive Psychology*, *60*, 218-240.

Christensen, I. P., & Huang, Y. L. (1979). The auditory tau effect and memory for pitch. *Perception & Psychophysics*, *26*, 489–494.

Cipolotti, L., Butterworth, B., & Denes, G. (1991). A specific deficit for numbers in a case of dense acalculia. *Brain*, *114*, 2619–2637.

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1954). Interdependence of temporal and auditory judgments. *Nature*, *174*, 642–644.

Cressie, N.A.C. (1993). *Statistics for Spatial Data, Revised Edition.* Hoboken, NJ: Wiley & Sons.

Crowder, R. G., & Neath, I. (1995). The influence of pitch on time perception in short melodies. *Music Perception: An Interdisciplinary Journal*, *12*, 379–386.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371-396.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219–250.

Dehaene, S., & Cohen (2007). Cultural recycling of cortical maps. *Neuron*, *56*, 384–398.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*, 1217–1220.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*, 970–974.

de Hevia, M.D., Izard, V., Coubart, A., Spelke, E.S., & Streri, A. (2014). Representations of space, time and number in neonates. *Proceedings of the National Academy of Sciences*, *111*(13), 4809-4813.

de Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, *21*, 653–660.

Diggle, P.J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns* (3rd ed.). Boca Raton, FL: CRC Press.

Dormal, V., Seron, X., & Pesenti, M. (2006). Numerosity-duration interference: A Stroop experiment. Acta Psychologica, 121, 109–124.

Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience*, *10*, 915–921.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341–354.

Dowling, W.J., & Fujitani, D.S. (1971). Contour, interval, & pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49*(2), 524.

Droit-Volet, S. (2010). Speeding up a master clock common to time, number, and length? *Behavioral Processes*, *85*, 126–134.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*, 131–136.

Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(24), 429-433.

Farnè, M. (1977). Brightness as an indicator of distance: Relative brightness per se or contrast with the background? Perception, 6, 287–293.

Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59–65.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, *65*(2-3), 263-297.

Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., & Verguts, T. (2010). Verbal-spatial and visuospatial coding of number–space interactions. *Journal of Experimental Psychology: General*, *139*(1), 180-190.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia. *Archives of Neurology and Psychiatry*, *44*, 398–408.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., and Fias, W. (2006). Numbers and space: a computational model of the SNARC effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 32–44.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York, NY: Appleton-Century-Crofts.

Ginsburg, N. (1991). Numerosity estimation as a function of stimulus organization. *Perception*, *20*, 681-686.

Goldstone, S., Lhamon, W. T., & Sechzer, J. (1978). Light intensity and judged duration. *Bulletin of the Psychonomic Society*, 12, 83–84.

Gottfried, A. W., Rose, S. A., & Bridger, W. H. (1977). Cross-modal transfer in human infants. *Child Development*, *48*, 118–123.

Gregory, R. L., & Wallace, J. G. (1963). Recovery from early blindness: A case study. *Experimental Psychology Society Monographs*, *2*, Cambridge: Heffers.

Harrington, D. L., Haaland, K. Y., & Knight, R. T. (1998). Cortical networks underlying mechanisms of time perception. *Journal of Neuroscience*, *18*, 1085–1095.

Henik, A., & Tzelgov, H. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition I*, *10*, 389–395.

Henry, M. J., & McAuley, J. D. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *Journal of Experimental Psychology: Human*

Perception and Performance, 35, 551–564.

Henry, M. J., McAuley, J. D., & Zaleha, M. (2009). Evaluation of an imputed pitch velocity model of the auditory tau effect. *Attention, Perception, & Psychophysics, 71*, 1399–1413.

Jacob, S.N., & Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *European Journal of Neuroscience*, *30*(7), 1432-1442.

Jacob, S.N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: the brain's code for proportions. *Trends in Cognitive Sciences*, *16*(3), 157-166.

January, D., & Kako, E. (2007). Re-evaluating evidence for the linguistic relativity hypothesis: Response to Boroditsky (2001). *Cognition*, *104*, 417–426.

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*, 983–997.

Körding, K.P., Beierholm, U., Ma, W.J., Quartz, S., Tenenbaum, J.B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS ONE*, *9*, e943.

Körding, K.P., & Wolpert, D.M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.

Kraemer, P. J., Brown, R. W., & Randall, C. K. (1995). Signal intensity and duration estimation in rats. *Behavioural Processes*, *34*, 265–268.

Kraemer, P. J., Randall, C. K., & Brown, R. W. (1997). The influence of stimulus attributes on duration of matching-to-sample in pigeons. *Animal Learning and Behavior*, *25*, 148–157.

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*. Chicago, IL: University of Chicago Press.

Lambrechts, A., Walsh, V., & van Wassenhove, V. (2013). Evidence accumulation in the magnitude system. *PLoS ONE*, *8*(12), e82122.

Leon, M. I., & Shadlen, M. N. (2003). Representation of time by neurons in the posterior parietal cortex of the macaque. *Neuron*, *38*, 317–327.

Levitin, D.J., & Rogers, S.E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*(1), 26-33.

Lewkowicz, D.J., & Minar, L. (2014). Infants are not sensitive to synaesthetic cross-modality correspondences: A comment on Walker et al. (2010). *Psychological Science*, *25*(3), 832-834.

Lourenco, S.F., & Longo, M.R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873-881.

Lourenco, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In S. Dehaene & E. Brannon (Eds.), *Space, time, and number in the brain: Searching for the foundations of mathematical thought* (pp. 225–244). Waltham, MA: Academic Press.

Lu, H., Chen, D. & Holyoak, K.J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*(3), 617-648.

Luce, R.D. (1990). "On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Psychological Review*, *97*(1), 66-77.

Luce, R.D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*(3), 520-532.

Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 384–394.

Marr, D. (1982). *Vision*: *A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: Freeman.

Marvin, E.W. (1997). Tonal/atonal: Cognitive strategies for recognizing transposed melodies. In Baker, Beach, & Bernard (Eds.), *Music Theory in Concept and Practice*. Rochester, NY: University of Rochester.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours: Extensions of a theory for contour. *Journal of Music Theory*, *31*, 225–267.

Maurer, D., Stager, C. L., & Mondloch, C. J. (1999). Cross-modal transfer of shape is difficult to demonstrate in one-month-olds. *Child Development*, *70*, 1047–1057.

McDermott, J.H., Lehr, A.J., & Oxenham, A.J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*(12), 1263-1271.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 320–334.

Meltzoff, A. N., & Borton, R. W. (1979). Intermodal matching by human neonates. *Nature*, *282*, 403–404.

Merritt, D. J., Casasanto, D., & Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition*, *117*, 191–202.

Moore, B.C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). Bingley, UK: Emerald Group.

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

Nagel, E., & Newman, J.R. (2001). *Gödel's Proof: Revised Edition*. D.R. Hofstader (Ed.). New York: NYU Press.

Nuñez, R., Doan, D., & Nikoulina, A. (2011). Squeezing, striking, and vocalizing: Is number representation fundamentally spatial? *Cognition*, *120*, 225–235.

Opfer, J. E., Thompson, C. A., & Furlong, E. E. (2010). The early development of spatial–numeric associations: Evidence from spatial and quantitative performance of preschoolers. *Developmental Science*, *13*, 761–771.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco: Morgan Kaufman.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Pariyadath, V., & Eagleman, D. (2007). The effect of predictability on subjective duration. *PLoS ONE*, *11*, e1264.

Peretz, I., & Hyde, K. L. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, *7*(8), 362–367.

Pfeuty, M., & Peretz, I. (2010). Abnormal pitch – time interference in congenital amusia: Evidence from an implicit test. *Attention, Perception, & Psychophysics*, 72, 763–774.

Piaget, J. (1952). *The Child's Conception of Number*. London, UK: Routledge.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York, NY: Basic Books.

Pinel, P., Piazza, M., Le Bihan, D., & Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron*, *41*, 1–20.

Plantinga, J., & Trainor, L. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*(1), 1-11.

Prince, J.B., Schmuckler, M.A., & Thompson, W.F. (2009). Cross-modal melodic contour similarity. *Canadian Acoustics*, *37*(1), 35-49.

Rammsayer, T. H., & Verner, M. (2014). The effect of nontemporal stimulus size on perceived duration as assessed by the method of reproduction. *Journal of Vision*, *14*(5):17, 1–10.

Rammsayer, T. H., & Verner, M. (2015). Larger visual stimuli are perceived to last longer from time to time: The internal clock is not affected by nontemporal visual stimulus size. *Journal of Vision*, *15*(3):5, 1-11.

Roberts, W. A. (1995). Simultaneous numerical and temporal processing in the pigeon. *Current Directions in Psychological Science*, *4*(2), 47–51.

Roitman, J. D., Brannon, E. M., Andrews, J. R., & Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychologica*, *124*, 296–318.

Roberts, W. A., Coughlin, R., & Roberts, S. (2000). Pigeons flexibly time or count on cue. *Psychological Science*, *11*(3), 218–222.

Roffler, S. K., & Butler, R. A. (1967). Localization of tonal stimuli in the vertical plane. *Journal of the Acoustical Society of America*, *43*, 1260–1266.

Rudel, R. G., & Teuber, H. L. (1964). Crossmodal transfer of shape discrimination by children. *Neuropsychologia*, *2*, 1–8.

Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*, 113–129.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin, 2*, 110–114.

Seitz, A.R., Kim, R., & Shams, L. (2006). Sound facilitates visual learning. *Current Biology*, *16*, 1422–1427.

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers: A cross-linguistic comparison of the SNARC effect. *Cognition*, *108*, 590–599.

Shams, L., Wozny, D.R., Kim, R., & Seitz, A. (2011). Influences of multisensory experience on subsequent unisensory processing. *Frontiers in Psychology*, *2*(264), 1-9.

Shepard, R.N. (1981). Psychological relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology*, *24,* 21-57.

Shigeno, S. (1986). The auditory tau and kappa effects for speech and nonspeech stimuli. *Perception & Psychophysics*, *40*, 9–19.

Srinivasan, M., & Carey, S. (2010). The long and the short of it: on the nature and origin of functional overlap between representations of space and time. *Cognition*, *116*(2), 217-241.

Stevens, S.S. (1975). *Psychophyisics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Stevens, J.C., Mack, J.D., & Stevens, S.S. (1960). Growth of sensation on seven continua as measured by force of handgrip. *Journal of Experimental Psychology*, *59*, 60-67.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Takayama, Y., Sugishita, M., Akiguchi, I., & Kimura, J. (1994). Isolated acalculia due to left parietal lesion. *Archives of Neurology*, *51*, 286–291.

Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., Rossetti, Y., et al. (2010). The amusic brain: Lost in music, but not in space. *PLoS ONE*, *5*(4), e10173.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, *71*, 259–264.

Trehub, S.E., Thorpe, L.A., & Morrongiello, B.A. (1987). Organizational processes in infants' perception of auditory patterns. *Child Development*, *58*(3), 741-749.

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences*, *104*(36), 14513–14518.

Vallentin, D., & Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. *Current Biology*, *18*, 1420–1425.

Vallentin, D., & Nieder, A. (2010). Representations of visual proportions in the primate posterior parietal and prefrontal cortices. *European Journal of Neuroscience*, *32*, 1380-1387.

van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7- to 9-year-olds. *Journal of Experimental Child Psychology*, *101*, 99–113.

van Opstal, P., & Verguts, T. (2013). Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Frontiers in Psychology*, *4*, Article 435.

Viarouge, A., & de Hevia, M.D. (2013). The role of numerical magnitude and order in the illusory perception of size and brightness. *Frontiers in Psychology*, *4*, 484.

Wagner, S., Winner, E., Cicchetti, D., & Gardner, H. (1981). “Metaphorical” mapping in human infants. *Child Development*, *52*, 728–731.

Walker, P., Bremner, J. G., Mason, U., Spring, J., Mattock, K., Slater, A., et al. (2010). Preverbal infants’ sensitivity to synaesthetic cross-modal correspondences. *Psychological Science*, *21*(1), 21–25.

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space, and quantity. *Trends in Cognitive Sciences*, *7*(11), 483-488.

Wilkie, D. M. (1987). Stimulus intensity affects pigeons' timing behavior: Implications for an internal clock model. *Animal Learning and Behavior*, *15*, 35–39.

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, *7*, 1–5.

Yates, M. J., Loetscher, T., & Nicholls, M. E. R. (2012). A generalized magnitude system for space, time, and quantity? A cautionary note. *Journal of Vision*, *12*(7):9, 1–7.

Yildirim, I., & Jacobs, R.A. (2012). A rational analysis of the acquisition of multisensory representations. *Cognitive Science*, *36*, 305-332.

Yildirim, I., & Jacobs, R., (2013). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition*, *126*(2), 135-148.

Zaytsev, J., & Chernyak, M. (2014). Fabric.js (Version 1.4.10) [Software]. Available from http://www.fabricjs.com/

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature*, *417*, 138–139.