Title
Giving Responses Dimension: Representational Shifts in Color Space and Event Segmentation Decisions in Physical Space Over Time

Author
Kelly, Laura Jane

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Giving Responses Dimension: Representational Shifts in Color Space and Event Segmentation Decisions in Physical Space Over Time

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Cognitive and Information Sciences

by

Laura Jane Kelly

Committee in charge:

Professor Evan Heit, Chair

Professor Rick Dale

Professor Michael Spivey

2018

The dissertation of Laura Jane Kelly is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Professor Rick Dale

Professor Michael Spivey

Professor Evan Heit, Chair

University of California, Merced

2018

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to foremost thank my committee chair, Evan Heit. Evan, your unflagging support as my advisor and collaborator has been an integral component in my success as a researcher. I am exceedingly grateful for all the opportunities you have offered me. Thank you for sharing your expertise as a scientist, both explicitly and by example. From experimental design, to critical theoretical analysis, to writing and presenting, you have challenged me to question my assumptions and biases, to express myself clearly, and to achieve more than I realized I could.

I would also like to thank my committee members Rick Dale and Michael Spivey. Rick, from your methodologies course my first year, to the speakers you have brought to campus, and to the questions you ask, you have regularly challenged me theoretically and technically. You've been a role model as a scientist and as a community member. Spivey, since taking your Foundations course, your ideas about continuity of mind and complex systems have influenced the way I view cognition, taking my initial leanings toward relativity from cognitive anthropology and giving me a language to talk about dynamics and flexibility. Thank you both.

I also want to thank the UC Merced Cognitive and Information Sciences Graduate Group. I am grateful for my time here, sharing in an intellectual community with the faculty, postdocs, and fellow graduate students with whom I have overlapped. Thank you for sharing your enthusiasm for cognitive science with me. Thank you especially to those of you who have become close friends over the years, sharing science, laughter, and life's ups and downs, particularly Sam, Bodo, Jordan, Shanna, Bryan, Chanita, Graham, Jeff, and Alex.

I also want to thank the undergraduate research assistants who stuck with me long-term: Maryam Nabizadeh, Aimee Nunez, and Elaine Ha. Your interest in science and your willingness to engage with my projects, contributing your own thoughts has been a highlight of my time at UC Merced.

The University of California, Merced, particularly the Graduate Division and the School of Social Sciences, Humanities and Arts, has been enormously supportive over the years, including generous financial support. In particular, the Spring 2018 Graduate Dean's Dissertation Fellowship has supported the completion of this dissertation.

Finally, thank you to my family and friends for your love and support. There are too many of you to name but a special thank you to my parents, Brian and Sue, and my siblings, JB and Ali. Also, to my pocket friends for your daily support.

—— ——

The text of Chapter 3 is a reprint of the article "Representational shifts made visible: movement away from the prototype in memory for hue" as it appeared in Frontiers in Psychology, which was co-authored by Evan Heit. The text of Chapter 4 is a reprint (with permission of the American Psychological Association) of the article "Recognition memory for hue: Prototypical bias and the role of labeling" as it appeared in the Journal of Experimental Psychology: Learning, Memory, and Cognition, which was co-authored by Evan Heit. The empirical work of Chapter 6 will be submitted for publication with co-author Evan Heit.

CURRICULUM VITAE

Laura Jane Kelly

lkelly@ucmerced.edu
LauraJaneKelly.com

EDUCATION

2018	Ph.D. in Cognitive and Information Sciences, University of California, Merced
2011	M.S. in Psychology, Lehigh University, Bethlehem, PA Cognitive Psychology
2009	B.A. in Anthropology, Lehigh University, Bethlehem, PA Highest Honors

PUBLICATIONS

Rotello, C. M., **Kelly, L. J.**, & Heit, E. (*Accepted*). The shape of ROC curves in shooter tasks: Implications for best practices in analysis. *Collabra*.

Kelly, L. J. & Heit, E (2017). Recognition memory for hue: Prototypical bias and the role of labeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 955-971. doi: 10.1037/xlm0000357

Kelly, L. J. & Heit, E (2014). Representational shifts made visible: movement away from the prototype in memory for hue. *Frontiers in Psychology 5*:796. doi: 10.3389/fpsyg.2014.00796

UNPUBLISHED MANUSCRIPTS

Rotello, C. M., Heit, E. & **Kelly, L. J.** Do modals identify better models? A comparison of signal detection and probabilistic models of inductive reasoning. *In Preparation.*

Kelly, L. J. Event perception and segmentation: Issues of representation and embodiment. *Ph.D. candidacy qualifying paper.*

Kelly, L. J. Is segmentation necessary? Evidence from language comprehension. *Ph.D. candidacy qualifying paper.*

CONFERENCE PROCEEDINGS

Kelly, L, J. & Heit, E. (2013). Representational shifts towards the prototype in memory for hue. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceeding of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

CONFERENCE PRESENTATIONS

Oral Presentations

Rotello, C. M., Heit, E. & **Kelly, L. J.** (2016). Do modals identify better models? A comparison of signal detection and probabilistic models of inductive reasoning. *57th annual meeting of the Psychonomic Society*, Boston, MA.

Rotello, C. M., Heit, E. & **Kelly, L. J.** (2016). Do modals identify better models? A comparison of signal detection and probabilistic models of inductive reasoning. New Approaches and Methods in Inductive Reasoning Symposium. *International Conference on Thinking*. Providence, RI.

Rotello, C. M., Heit, E. & **Kelly, L. J.** (2015). Accuracy differences for congruent and incongruent belief bias problems reflect response bias. *56th annual meeting of the Psychonomic Society*. Chicago, IL.

Kelly, L. J. & Heit, E. (2014). Direction of representational shifts in memory for hue is not fixed. *7th Embodied and Situated Language Processing Conference*. Rotterdam, Netherlands.

Kelly, L. J. & Malt, B. C. (2011). Color language's effect on perception: The activation of linguistic categories. Paper presented at *The 82nd Annual Meeting of the Eastern Psychological Association.* Cambridge, MA.

Poster Presentations

Heit, E., **Kelly, L. J.**, & Rotello, C. M. (2016) ROCs Show That Shooter Bias is Response Bias. Poster presented at *57th annual meeting of the Psychonomic Society*, Boston, MA.

Kelly, L. J. & Heit, E. (2014). Labels Focus Attention: Implications for Language Relativity. Poster presented at *Conceptual Structure, Discourse, and Language 12*, Santa Barbara, CA.

Kelly, L. J. & Heit, E. (2013). Representational shifts towards the prototype in memory for hue. Poster presented at the *35th Annual Conference of the Cognitive Science Society*, Berlin, Germany.

Kelly, L. J. & Heit, E. (2012). Representational shifts in memory for hue. Poster presented at the *53rd annual meeting of the Psychonomic Society*, Minneapolis, MN.

INVITED TALKS

2018	Navy Center for Applied Research in Artificial Intelligence Naval Research Laboratory, Washington D.C.

FELLOWSHIPS, SCHOLARSHIPS, AND PRIZES

2018	Graduate Dean's Dissertation Fellowship, UC Merced
2016	Student Scholarship to Trieste Encounters on Cognitive Science, Language Learning
2013 – 2018	Cognitive and Information Sciences Travel Fellowship, UC Merced
2013 – 2016	Cognitive and Information Sciences Summer Fellowship, UC Merced
2012	Graduate Research Council Summer Fellowship, UC Merced
2012	Graduate Division General Fellowship, UC Merced
2011 – 2015	Chancellor's Graduate Fellowship, UC Merced

2009 – 2010	President's Scholar, Lehigh University
2009	Donald T. Campbell Social Science Research Prize, Lehigh University Second Place Undergraduate Prize for *Hierarchies of the Senses*
2005 – 2009	Dean's Scholarship, Lehigh University

HONORS AND AWARDS

2009	Department of Sociology and Anthropology, Lehigh University Departmental Honors
2007 – 2009	Eckardt College Scholar, Lehigh University

RESEARCH EXPERIENCE AND POSITIONS

2013 – 2017	Graduate Student Researcher – Heit Lab, with Dr. Evan Heit
2009 – 2011	Master's Thesis – Psychology Master's Program, with Dr. Barbara Malt *Investigating the effect of overt labeling on lateralized categorical perception for color.*
Spring 2010	Research Assistant – Language Production Lab, with Dr. Padraig O'Seaghdha
2008 – 2009	Senior Thesis – Anthropology Major, with Dr. John Gatewood *Hierarchies of the Senses: Investigating Color, Texture and Scent in American English.*

ADDITIONAL TRAINING

June 2013	APA Advanced Training Institute: Exploratory Data Mining in Behavioral Research. University of California, Davis
June 2012	APA Advanced Training Institute: Nonlinear Methods for Psychological Science. University of Cincinnati.
Spring 2008	School for International Training, Brattleboro, VT SIT Samoa: Pacific Island Studies

TEACHING EXPERIENCE

Instructor of Record, University of California, Merced

Summer 2017	Introduction to Cognitive Science

Teaching Assistant, University of California, Merced

Fall 2017	Contemporary Moral Problems
Spring 2013	Cognitive Neuroscience
Fall 2012	Introduction to Cognitive Science
Spring 2012	Perception
Fall 2011	Judgment and Decision Making

Teaching Assistant, Lehigh University

Spring 2011	Child Development
Fall 2010	Statistical Analysis of Behavioral Data

SERVICE

2014-2017	Cognitive Science Internal Graduate Student Representative
2013-2017	Cognitive Science Graduate Student Representative to the UCM Graduate Student Association
2015	Co-Organizer of Prospective Student Visit Weekend

Ad-hoc Reviewer for: *Frontiers in Psychology, Journal of Experimental Psychology: General*

PRESS

Particularly Exciting Experiments in Psychology: Labels and Memory
https://www.apa.org/pubs/highlights/peeps/issue-91.aspx

Codecademy Stories: Study the human mind – with Python
http://www.codecademy.com/stories/study-the-human-mind-with-python

Abstract

Contextualized through the lens of field-wide debates of representation theories and embodiment theories which dictate assumptions about the nature of cognition, this dissertation presents research on two cognitive phenomena, representational shifts and event segmentation. Both phenomena intersect traditional divisions of cognitive processing spanning perception, attention, memory, and decision-making. A fundamental challenge in the behavioral cognitive sciences is designing experiments and contexts that elicit behavior that reflects cognitive processes. The title phrase, "Giving Responses Dimension," reflects the strategy employed across the empirical chapters to elicit more informative behavior through dimensional responses, either by projecting discrete responses into abstract conceptual color space in the case of representational shifts, or by movement dynamics of responses extending over physical space with hand/mouse movements.

The representational shift experiments show contextual influence on memory bias. In a rapid study-to-test paradigm, memory for hue is biased away from the basic color category of that hue. At a longer delay, memory is biased toward the color category of the hue. Labeling the category of a hue mitigates the size of the bias toward the category. Being forewarned of the memory test produced the same reduced bias. In both situations, the experimental context creates cues indicating which aspects of perceptual information may be most useful in the future. Memory was then biased to retain those aspects of the perceived items expected to be useful.

The event segmentation experiments examine the decisions that go into responding to the event segmentation task. When asked to segment a video by marking off the end of "natural and meaningful activity units" of different sizes, people have some degree of consistency with themselves and others in terms of response timing. Quantifying this consistency has been the major mode of analyzing results from this task in the literature. By adding continuous response measures on top of or instead of the discrete responses, the decisions underlying those responses can be examined. The changes to the paradigm do not alter the basic patterns of individual trial to group segmentation consistency. When asked to predict the end of an activity unit, segmentation responses are predicted by the responder over a third of the time. When the segmentation response itself is continuous, participants employ diverse response strategies as well as

exhibit sub-response threshold movements toward a segmentation response in addition to the full segmentation responses. These novel measures, reflecting decisions or reasoning about decisions, provide new insight into the behavior underlying a task with substantial importance for theories of event cognition.

This dissertation, *Giving Responses Dimension: Representational Shifts in Color Space and Event Segmentation Decisions in Physical Space Over Time*, is submitted by Laura J. Kelly in 2018 in partial fulfillment of the degree Doctor of Philosophy in Cognitive and Information Sciences at the University of California, Merced, under the guidance of dissertation committee chair Evan Heit.

Chapter 1

Introduction

1.1 Beyond Object Metaphors

Does the language researchers use to discuss cognition bias the thoughts the researchers have about how cognition functions? This is a cognitive scientist specific version of the Whorfian hypothesis (Whorf, 1956), the idea that a speaker's habitual language affects his or her habitual thought processes. More broadly than just language, the idea of habitual experience creating habitual bias in perception, memory, and other developmentally plastic processes in each unique cognizer has been a guiding principle in how I approach my own research.

Cognitive scientists are as subject to the influence of our own past experiences as our participants are. Only carefully examining the assumptions that go into scientific methods and theories allow us to push past incorrect, but ingrained, theories. Churchland (1981) goes so far to suggest progress in the cognitive sciences depends on researchers divorcing themselves from folk psychology terms. The basic divisions of cognition such as perception, attention, memory, reasoning, and language capacity are divisions that were drawn and cemented in the lexicon before we developed a scientific understanding of underlying mechanisms. Developing a unique scientific language could codify a change in the default conception of cognition and its functions. I will not propose a new language for cognitive science; this is a change that is slow and ongoing. In this dissertation, I will be questioning assumptions many of which default to treating cognition as divided into processes along the tradition division lines just outlined. In the present empirical work, these assumptions are investigated through the use of a finer-grain of measurement than previous research on the phenomena of interest, representational shifts and event segmentation. Cognitive science is currently evolving to be less the study of isolated components of a broader cognitive system and instead to be more focused on interactions between the traditionally defined processes. Referring to processes as nouns, e.g. the perception or the memory, gives the impression that there is a thing, a delineable item that is the cognitive process or result thereof. But what would that thing be? A memory is not a piece of mental paper that can be read, filed, retrieved, and re-read. A memory is a complex cascade of neural activation patterns unfolding over time. The thing we call a memory is the complex relationship between neural connections, transient firing activity, and millions of other memories making use of the same neural hardware.

By using traditional component centered theoretical language, researchers may be biasing themselves to be too narrow in their assessment of the range of possibilities about the phenomena they study. If a researcher considers a memory to be an object-like representation in the mind, then there are certain properties of objects that will be assigned to it, the memory, such as molecular cohesion and physical permanence that are not necessarily part of a cognitive object. If memory is instead thought of as a process, a cascade of neural firing in complex sequences and with complex consequences, the realm of measurable cause and effect around memory changes. From this perspective, drawing a line around some aspect of the larger cognitive system may be necessary for limiting the scope of investigation but such a defined section of cognition is not interesting in and of itself without the context of the rest of cognition. The line between perception and memory or perception and action is fuzzy at best and in some circumstances may not be a relevant distinction.

The language of cognitive science is, however, currently very categorical in nature. If a researcher wants to be in conversation with other researchers, they need to use the common parlance. As much as cognition may not be as discrete as object metaphors imply, we need to discuss aspects of a cognitive system in some way. Therefore, the in this dissertation are frequently discussed using the terms assigned to them by other researchers. These terms are often object metaphors as is the case with the two theoretical phenomena under empirical study, representational shifts and event segmentation.

1.2 Perception and Memory

Perception is a process of taking sensory input and making sense of it in some way - combining individual signals from individual receptor cells into coherent patterns that hold meaning for the perceiver. For example, vision begins with electromagnetic waves being detected by cells in the eyes that are sensitive to particular wave amplitudes. Inferred from the signals sent from these cells, humans perceive color, shape, orientation, and more. As perceptual processing continues, the combination of co-occurring low-level inferences produce more complicated higher-level inferences such as motion and depth. Perception in this way can be viewed as a dimension reduction process on new information.

Memory is the information signal that has been transduced into neural activation patterns. Memory has traditionally been broken down taxonomically into different forms and stages. Initial signals are considered sensory memory. After perceptual processing, signals are in working memory until they are stored in long-term memory of which there are many types. Any further processing that happens on the signals are considered memory processes, for instance, memory maintenance, storage, and retrieval.

Perception and memory are integral to each other. In the way I've presented it above, perception is a process that results in memory. But during perceptual processing, the signals are already sensory memory. Further, what is the difference between experiencing the external world vs. experiencing the reactivation of long-term memories? Is retrieving memories not perception of a sort as well? In the case of memory retrieval, the input to working memory processing is high dimensional neural activations rather than exclusively low-level sensory input, but the processing in both cases involves finding meaningful patterns from the input. In fact, it is well established that reactivated memory is malleable and subject to change, resulting in altered memories later mistaken as memory of the original perception (Loftus & Palmer, 1974; Nadel, Hupbach, Gomez, & Newman-Smith, 2012). This is evidence of new pattern identification with input from long-term memory. Yet, likely because of traditional delineations of cognitive stages and processes involving largely independent memory retrieval processes, input from the senses or from memory are treated as substantively different.

The empirical work in this dissertation falls in this area of joint perception/memory. With representational shifts, in Chapters 3 and 4, color hues are perceived, categorical labels are perceived and generated, category knowledge is retrieved, and, finally, newly stored memory of the hues is tested for accuracy. With event segmentation, in Chapter 6, ongoing visual and auditory stimuli, videos of simple activities, are perceived and broken up into meaningful parts by participants. This segmentation of the videos depends on both the newly perceived information and the perceiver's existing knowledge of how the activities being depicted are usually conducted. These empirical investigations span multiple categories of cognitive processing that have traditionally been studied separately.

1.3 Dissertation Roadmap

1.3.1 Theoretical Issues of Representation and Embodiment

Chapter 2 presents two core theoretical issues within the cognitive sciences: representation and embodiment. The diversity of views on these topics will be reviewed in detail. Representation is retaining information in mental symbols that represents, or stands in place of, aspects of the external world. Many researchers think it is a basic and necessary part of any cognitive theory while others question the concept's utility. Embodiment is the extent to which a cognizer's physical implementation - its brain, its body, perhaps even its environment - affects its cognition. The chapter reviews the three major types of embodiment theories according to Lawrence Shapiro's taxonomy: conceptualization, replacement, and constitution embodiment (Shapiro, 2011). Additionally, there are researchers who reject embodiment completely. The chapter contains an integrative discussion of how some theoretical stances on each issue are complementary and others are mutually exclusive. Finally, the chapter concludes with a review of a currently active debate over whether embodied representation constitutes a paradigm shift in the Kuhnian (Kuhn, 1962) sense.

1.3.2 Representational Shifts

At any moment, we are perceiving a range of sensory information - visual, auditory, tactile, etc. Some sensory information is retained in memory while other information is forgotten. As mentioned above, it is well known that human memory is often not accurate to objective records of past experiences (Loftus & Palmer, 1974; Nadel et al., 2012). Quantifying how memory is biased in comparison to original experience can provide insight into how information is distorted by the human cognitive system. Chapters 3 and 4 pursue this objective, reporting empirical investigations of the interaction between visual perception and online labeling through experimental testing of memory accuracy. Specifically, the chapters test claims of the Representational Shift Hypothesis (Lupyan, 2008):

> *… when category labels are activated, they produce top-down feedback that activates visual features stored with the category on previous occasions. The features activated by top-down processing become coactive with features activated through bottom-up processing. As activation patterns continue to cycle, the active visual features settle on those that are consistent with both bottom-up input from the exemplar and top-down input from the category (McClelland & Rumelhart, 1981). This results in a mismatch between the stored representation of the studied item and the retrieval cue (the studied item presented during the testing phase), which in turn should produce more 'new' responses for old items (i.e., a lower hit rate). When category labels are less active, or when the top-down activity is interfered with, the representation of the visual input is closer to the bottom-up information presented than when labels produce top-down inferences about the category's features. This results in a closer match between the studied item and the item presented at test (the retrieval cue), thus resulting in a higher proportion of hits (p. 349).*

The hypothesis explicitly claims that memory for sensory experience is altered by the concurrent labeling of a category during the original sensory experience. The difference occurs during the encoding of the memory trace. According to the hypothesis, a trace should be more biased away from the original percept if the category was concurrently labeled than if it was not. The bias is relative to category structure. These representational shifts are suggested to be in the direction of being more representative, or more typical, of the category.

The empirical work presented in Chapters 3 and 4 use experimental designs that aimed to collect more graded information than had been collected in the original experiments. Within the well quantified dimension of color hue, a pattern of multiple lures that could be confused for the target hues were selected. The distribution of sensitivity scores, a measure of confusability of lures with the target, along a typicality gradient, the directional distance in hue from lure to target relative to better examples of the color category, allows for inferences about the direction and strength of memory test performance shifts from a balanced error pattern. Additionally, Chapter 3 addresses the encoding claim, finding little evidence of an effect of labeling on sensitivity by typicality soon after initial encoding. Chapter 4 addresses the labeling claim at a longer timescale finding an effect. However, the effect is one of less bias when hues were labeled rather than more.

Chapter 3 is published in Frontiers in Psychology (Kelly & Heit, 2014). Chapter 4 is published in the Journal of Experimental Psychology: Learning Memory and Cognition (Kelly & Heit, 2017). Below are the article abstracts.

1.3.2.1 Representational Shifts Made Visible: Movement Away from the Prototype in Memory for Hue[1]

In four experiments, a total of 205 participants studied individual color patches and were given an old-new recognition test after a brief retention interval (0.5 or 5.0 s). The pattern of hue sensitivity (d′) revealed hue memory shifting away from the prototype of the hue's basic color category. The shifts demonstrate that hue memory is influenced by categorization early in processing. The shifts did not depend on intentional categorization; the shifts were found even when participants made preference ratings at encoding rather than labeling judgments. Overall, we found that categorization and memory are deeply intertwined from perception onward. We discuss the impact of the results on theories of memory and categorization, including the effects of category labels on memory (e.g., Lupyan, 2008). We also put forward the hypothesis that atypical shifts in hue are related to atypical shifts that have previously been observed in face recognition (Rhodes et al., 1987).

1.3.2.2 Recognition Memory for Hue: Prototypical Bias and the Role of Labeling[2]

How does the concurrent use of language affect perception and memory for exemplars? Labels cue more general category information than a specific exemplar. Applying labels can affect the resulting memory for an exemplar. Here three alternative hypotheses are proposed for the role of labeling an exemplar at encoding: (1) labels distort memory towards the label prototype, (2) labels guide the level of specificity needed in the current context, and (3) labels direct attention to the label's referent among all possible features within a visual scene. University students were shown hues on object silhouettes that they either labeled with basic color categories, made preference judgments about, or indicated the animacy of its category. Experiments 1 and 2 established that there are response shifts toward the category prototype regardless of labeling showing a pervasive

[1] The official citation that should be used in referencing this material is Kelly, L. J. & Heit, E. (2014). Representational shifts made visible: Movement away from the prototype in memory for hue. *Frontiers in Psychology, 5*.

[2] The official citation that should be used in referencing this material is Kelly, L. J. & Heit, E. (2017) Recognition memory for hue: Prototypical bias and the role of labeling. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43, 955-971.

influence of category knowledge on response bias. They also established an effect of labeling whereby labeling decreases the magnitude of shifts. Experiment 3 and 4 investigated the uniqueness and necessity of language in causing the decreased shift – neither of which proved to be the case. Overall, category-relative bias was pervasive and labeling appears to direct attention to the feature resulting in less biased memory. The results highlight that the context at encoding affects how memory is formed.

1.3.3 Event Perception

Event perception is the extraction of meaningful structure from changes in stimuli over time. One way to quantify this structure is to segment ongoing activity into meaningful parts. A prominent account of this process is presented in the Event Segmentation Theory (Zacks, Speer, Swallow, Braver, & Reynolds, 2007), a theory specifying the structure of processing streams relating new sensory input to existing knowledge and predictions of that sensory input on the part of the perceiver as events are experienced in real time. While the theory will be presented in detail in Chapter 5 "Event Segmentation Theory and Empirical Evidence Review," below are the major implications of the theory in the authors' own words:

> *The principal novel features of EST are that event models maintain stable representations of "what is happening now" and are updated based on transient increases in perceptual prediction error. The theory has several implications for perception and cognition:*
>
> 1. *Most important, the theory implies that the segmentation of ongoing activity into discrete events is a spontaneous concomitant of ongoing perception and does not require conscious attention.*
> 2. *Event segmentation is a mechanism of cognitive control. The gating mechanism resets event models and thus is the means by which the cognitive system exerts control over the disposition of processing resources and the updating of working memory.*
> 3. *Event segmentation happens simultaneously on multiple timescales, though an observer may attend to a particular timescale.*
> 4. *Event segmentation incorporates information from multiple senses. This results from the fact that the mechanisms in EST are general across sensory modalities and incorporate information from multiple modalities.*
> 5. *Event segmentation depends on change. When the world is static, prediction is easy.*
> 6. *Event segmentation depends on prior knowledge. Event models are constructed through the interaction of sensory input with stored knowledge (including the knowledge stored in event schemata).*
>
> *(Zacks et al., 2007, p. 277)*

This is a wide-reaching theory. The theory's "novel features" of stable representations of current experience and discrete updating of those representations based on prediction error are based on a strongly representational theory of cognition. This theory is incompatible with less strong views of representation. The first two implications are bold claims. Human sensory experience is continuous, only limited by collective firing rates of sensory receptor cells. As mentioned above, low-level perception is a process of dimension reduction, extracting patterns from sensory input with each reduced pattern feeding into another layer of reduction. The implications suggest the process of reduction ends with comparison to an event model and the signal only continues on to contribute to higher-order processing if sufficiently different from the existing event model. The theory suggests that the world is perceived in component parts and these discretized scenes are what become memory.

Chapter 5 reviews the EST in more depth, as well as the empirical evidence for ongoing segmentation, hierarchical structure, and memory effects. Each of the theory's specific claims and corresponding evidence on these topics are evaluated relative to the theoretical dimensions of representation and embodiment presented in Chapter 2.

Chapter 6 is an empirical investigation of the event segmentation task using continuous response measures to collect more information about the segmentation decisions made while performing the task. Specifically, participants' predictions of event boundaries are elicited in Experiment 1 using a continuous response slider and the action dynamics of the segmentation decision are collected in Experiment 2 through tracking segmentation responses produced by dragging a slider across a track. The slider represents a novel use of mouse tracking in a single dimension rather than the more frequently utilized two dimensions. Mouse trajectories over space and time can be used to infer information about a participants' decision-making from metrics such as velocity, acceleration, pauses, and reversals (Spivey, Grosjean, & Knoblich, 2005; Dale, Kehoe, & Spivey, 2007).

Everyday perception is dynamic and extended in time. While the approach to measuring representational shifts employed in Chapters 3 and 4 is more graded and less discrete than a target and a single lure used in previous research, the basic phenomenon under study is memory for a single feature, hue, of static stimuli. The measurements used to get a graded sensitivity distribution were forced choices between discrete options rather than direct graded responses. Following the same principle of gathering a finer-grained dependent measure, the experimental design implemented in Chapter 6 includes dynamic stimuli and a continuous response format, allowing for directly measured graded responses, increasing ecological validity.

A future version of this work will be submitted for publication with my co-author Evan Heit. The following section is an abstract for the empirical chapter.

1.3.3.1 Event Segmentation Decisions

Does the event segmentation task reflect automatic segmentation in perceptual processing? When asked to segment a video by marking off the end of "natural and meaningful activity units" of different sizes, people have some degree of consistency with themselves and others in terms of response timing. Quantifying this consistency has been the major mode of analyzing results from this task in the literature. By adding continuous response measures on top of or instead of the discrete responses, the decisions underlying those responses can be examined. The changes to the paradigm do not alter the basic patterns of individual trial to group segmentation consistency. When asked to predict the end of an activity unit, over a third of segmentation responses are reported as expected by the responder. When the segmentation response itself is continuous, participants used diverse strategies to transverse the response slider and exhibit sub-response threshold movements toward the response location in addition to full segmentation responses. Together, the two experiments show that the event segmentation task is reasoning task, not an exclusively perceptual task. These novel measures, reflecting decisions or reasoning about decisions, provide new insight into the behavior underlying a task with substantial importance for theories of event cognition.

1.3.4 General Discussion

Finally, in Chapter 7, I will review the material presented in Chapters 2-6 and integrate across them. In particular, I will return to the theoretical issues of representation and embodiment raised in Chapter 2 and reflect on the empirical studies in Chapters 3, 4, and 6 through the lens of these broad debates in cognitive science. I will also make connections from the empirical results to object metaphors, pragmatic context, and event cognition, as well as discuss why these are important issues in cognitive science. The theoretical importance of giving responses dimension will be reiterated and the conclusion will contain my own views on representation and embodiment informed by the work of this dissertation.

Chapter 2

Theoretical Issues of Representation and Embodiment

2.1 Representation

Representation is a much-contested topic centering around the questions: What are the cognitive objects in the mind? Do we create and store mental models of the world we interact with? If we do, how do they come to be? If we do not, how do we have detailed memory or language? There are contrasting problems of not knowing how symbols or symbol-like mental objects can come to be and not knowing how higher order complex cognition could happen without something symbol-like. How can we understand the world without having a mental dictionary of meaningful objects, events, and processes with which to compare new experiences? Yet, how do we have or build such a dictionary? These questions have been being discussed and examined since psychology emerged as a field of study. In fact, these debates date at least back to ancient Greek philosophy with Plato’s ideals: all items in the world are imperfect exemplars of a category and the true perfect item – an ideal – is available only on a spiritual plane (Ross, 1951). The debate continues in recent literature with some theorists suggesting higher-order cognitive processes would not be possible without representations (Clark & Toribio, 1994; Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016) and others suggesting they are a crutch preventing more nuanced theorizing (Barrett, 2011; Chemero, 2009; Wilson & Golonka, 2013).

2.1.1 Definition

Mental representations are traditionally used as “causally potent information carrying vehicles” in theories of cognition (Chemero, 2009, p.50). They are an abstract concept with different researchers defining them in different ways for each cognitive domain, such as active representation for working memory, more permanent storage representations in long-term memory, plan representations for speaking or acting in the immediate future, etc. With so many different cognitive ‘things’ getting the same name, talking about the general idea of representation can be difficult. Connell & Lynott (2014) supply a useful taxonomic breakdown to divide these ideas of representation into the action and the object: representation and representing refer to current activation and processing, while concepts are the more stable long-term connectivity patterns. Regardless of how the terms are specifically applied by particular researchers, representation is mental and internal. These basic characteristics are the features on which I will focus the current discussion.

2.1.2 History

The major shifts in psychological mainstream theory center around shifts in majority thought about representation. Edward B. Titchener, Wilhelm Wundt, as well as other introspectionists viewed the primary goal of psychology as understanding human conscious experience (Baars, 1986). There is abundant meaningful content within conscious experience and introspectionists believed careful examination of conscious thought by highly trained psychologists would be fruitful for understanding that thought content. Introspectionists focused on a sub-set of representations that can be investigated consciously. However, since self-examination is not conducive to reproducible empirical science and psychology became more serious about being a science, the first major shift in psychological theories was away from introspectionism toward behaviorism.

Behaviorism is characterized by careful, scientifically rigorous examination of external behavior. Internal, mental activity is discounted and believed unsuitable for study because it cannot be directly measured (Baars, 1986). Behaviorism research focuses on topics such as signal-response theory and conditioning which do not make claims about what is occurring within an organism, instead documenting the input/output relationships and the change in those relationships based on observable factors such as training.

After decades of behaviorism as the predominant theory in cognitive psychology, many researchers believed there was more going on than input and output to the system and began to study the 'more' as part of psychology. According to Baars (1986), "psychologists did not speak of 'mental representation' at first, but of 'memory'; not of 'consciousness,' but of 'selective attention'; not of 'the organization of meaning,' but of 'semantic features'" (p. 142). Any mental processing that occurs between perceptual input and action-based output was of key interest to these researchers who were not eliciting enough explanatory power from the input and output patterns alone. The subsequent cognitive revolution opened up theoretical discussion by elevating cognitive processes to be the main objects of psychological study. Behavioral data continued to be collected, but the objective was no longer just predicting output. It was also inferring what cognitive processes were happening inside an organism. These processes were generally considered to be representational with sensory input resulting in mental information - information that stands for aspects of the external world that can then be computed upon by the brain.

2.1.3 The Current Situation

Much of cognitive science is still centered on processing within the brain. Extensive methods and theories have been build around understanding the contents of brain-based cognition – generally representations and concepts – as well as the processes those representations go through such as perception, working memory, long-term memory, and so on. The major cognitive science and cognitive psychology academic journals currently predominantly publish articles communicating research of this type.

However, not all current empirical cognitive science is exclusively interested in brain-internal cognitive processing. In addition to some holdover introspectionism and behaviorism, there is a branch of psychology – ecological psychology – that views cognition to be of the complete agent-environment system rather than just within a cognitive agent. An agent-environment system is comprised of an agent, its environment, and, most importantly, the relationships between them including action opportunities. Therefore, ecological psychologists view the object of study as not just the processing within an organism's brain. They are equally, if not more, concerned with the organism's body, with its capacities for sensory input and action output, as well as the way those capacities interface with the organism's environment. Radical Embodied Cognitive Science (RECS), a specific philosophical stance incorporating many tenets of ecological psychology, ascribes to this understanding of cognition as beyond the brain with the additional stance of being explicitly anti-representational (Chemero, 2009; Wilson & Golonka, 2013). RECS views cognition as consisting of continuous dynamic processes rather than a series of discrete states.

2.1.4 Objects vs. Processes

Representations are conceptually tractable as mental objects. Humans live in a world of physical objects and have extensive experience interacting with them. Drawing a metaphor between the manipulation of physical objects and the manipulation of mental objects has been fruitful in studying cognition for decades. Philosophically, however, there are a number of intractable issues with representation. The symbol-grounding problem (Searle, 1980; Harnad, 1987) is the philosophical problem of connecting the physical world with a theorized symbolic mental world – how does a mental symbol come to stand for something in the physical world? Universal grammar (Chomsky, 1957) is a famous example of considering the building blocks of symbolic understanding – in this case of linguistic syntax – to be innate. By pushing the advent of primitive symbols within an individual to biology rather than learning, the grounding issue is side-stepped. Another way to address the symbol-grounding problem is to view symbols as being learned and grounded in sensorimotor processes (Barsalou, Simmons, Barbey, & Wilson, 2003). To illustrate this hypothesis concretely, grounded cognition suggests that when we think of a Red Delicious apple, we do not just think of the association with the concept red but we actively simulate the shade of red we associate with that particular variety of apple in the same brain networks as we process sensory perception of the color. Both of these solutions are part of representational accounts of cognition. Neither biological universals nor grounded representations are universally accepted as a solution. Both solutions agree, however, that the symbol-grounding problem is in fact a problem researchers need to solve.

Ecological psychologists instead suggest that mental objects are just a metaphor - there are no mental objects. While relating cognition to the physical world makes it more comprehensible in some ways, it also obscures the continuous and process driven reality of the system. For ecological psychologists, rather than cognizers developing mental representations that hold meaning, meaning instead comes from mutuality relations – the causal dependencies between components of a system. The meaning in a mechanical clock comes out of causal dependencies amongst the spring, gears, and other internal machinery that result in even movement of the hands of the clock over time. Likewise, according to ecological psychology, the meaning of an action comes from the changes in opportunities for future action – affordances – for the organism. The continuously changing relational properties of a system are the informational content carrying 'vehicles' but they are not object-like and they are not exclusively internal.

Contemporary cognitive science rarely takes a fully symbolic-object stance on representation. Traditional representations have what Clark and Toribio (1994) refer to as "quasi-linguistic combinatorial structure" (p. 401), which are symbolic, discrete, and have algorithmic rules for processing. However, Clark and Toribio point out that even less object-like distributed patterns in connectionist networks are still representations. Not being able to point to a simple mechanism such as grandmother cells – single cells firing selectively in response to a particular stimulus such as one's grandmother or a particular celebrity's face (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) – as concrete, symbolic internal representations does not mean that the concept of internal representation is wrong. Within networks, patterns of activation can compositionally represent concepts less transparently and discretely than the cell storage idea. Clark and Toribio defend internal representation from the place of assuming they exist until non-representational researchers can explain how 'representation-hungry tasks' can be accomplished without representation. An example of a representation-hungry task would be selecting the valuable items from an array (Clark & Toribio, 1994, p. 419-420). No perceptual cue encodes valuableness in the environment. Non-environment sourced knowledge has to be used to accomplish the task. A valuableness concept needs to be applied to the perceptual information. Clark and Toribio challenge anti-representationalists to account for this ability without that internal concept.

There has been empirical success for anti-representational theorists in what Clark and Toribio view as non-representation-hungry tasks. Robots have been created which can navigate complex environments successfully, and therefore display complex behavioral sequences, with just a few guiding equations (Brooks, 1991a, 1991b; Brooks, Breazeal, Marjanovic, Scassellait, & Williamson, 1999). This is an example of non-representationally driven behavior. Another example is the expert catching of a fly ball. The perceptual input and time limitations make it impossible to account for the ability using the representational and computational method of representing the key variables such as wind speed, angle of accent, etc., and calculating likely trajectories of the ball as well as the catcher's own capabilities in terms of speed and reach. Instead, with practice, a baseball player can learn to simply manipulate their own speed and heading to keep the perceptual experience of the ball looking a certain way (Wilson & Golonka, 2013). This is viewed as non-representational as all the information needed in the moment is directly perceptually available.

2.1.5 Conclusion

The majority of current researchers likely at least tacitly agree with a representational stance of some sort. The debate over what those representations would look like in the brain is far from settled. As presented above, a vocal group of researchers continue to question the existence of representations at all.

2.2 Embodiment

As introduced above, the content of cognition is debated on another dimension beyond representation: the extent to which cognition is abstract or embodied. Embodied cognition – the idea that the biology and experience of the agent beyond its brain impact the contents of that agent's cognition – is a relatively new theoretical stance. Shapiro (2011) reviews the state of embodied cognition identifying 3 main hypotheses: conceptualization, replacement, and constitution. These hypotheses each ascribe to some version of the body and/or environment beyond the brain influencing cognition. Each of these hypotheses also have a stance on representation which guides the theorized ways that the body and environment can influence cognition.

The conceptualization hypothesis suggests representations are affected by the congizer's physical form. It subsumes areas of research and theories dealing with how mental representations are affected by experience within a particular body and/or in a particular environment. An example of a research topic with theories of the conceptualization type is the area of language and thought which is concerned with whether, and how, the language an individual speaks influences their other cognitive processes and abilities. An embodied stance would be that the bodies with which we experience the world lead us to certain metaphors through which we scaffold our understanding of the world. This conception of embodied cognition is most similar to standard cognitive science, viewing cognition still to be mainly brain-based with representations being affected by the body and environment.

The replacement hypothesis encompasses anti-representational approaches, such as Radical Embodied Cognitive Science (Chemero, 2009) mentioned above. The replacement hypothesis critiques standard cognitive science as having been treating the brain as if it is fundamentally different from the rest of world. The brain is part of the body and the world. It should therefore be governed by the same principles. One path to rectifying this historical divide is describing cognition in terms of dynamical systems (Spivey, 2007); the same dynamical systems that can be used to describe any other physical processes. Brain networks may be specialized in their structure and capabilities, but ultimately they are physical processes.

The third hypothesis, what Shapiro calls the constitution hypothesis, expands the domain of what is considered the mind and cognition. Rather than the mind being the product of the brain alone, there are such integral interactions with the body and environment that those components are most usefully thought of as also constituting the mind. When an organism moves to alter or enhance visual input, the movement is part of visual perception and a form of cognition (O'Regan & Noë, 2001); when a person takes notes to be viewed later, the notes are external memory and a form of cognition (Clark & Chalmers, 1998). The argument rests on how central non-brain components are to cognitive processes. If a process could not occur without those parts, then this hypothesis would count them as cognitive. In this conception of embodied cognition, cognition is extended beyond the brain and even beyond the body. Representation is still a core component of cognition but is not exclusively mental. Defining the scope of the domain of cognition – the brain; the brain and the body; or the brain, the body, and the environment – is a key theoretical discussion.

These three subsets of embodied cognition span a wide range of theories, many of which are not compatible with each other. What all versions of embodied cognition have in common is they fundamentally challenge the idea that the brain is like a computer. The computer metaphor of the mind (Newell & Simon, 1976) is the idea that brains are computational and have parts which mimic computer components like a central processor (executive functions), RAM (working memory), disk storage (long term memory), and a set of peripherals such as keyboards, mouses, and screens (sensory and motor systems) that receive input or output the computationally transformed information. The computer metaphor has dominated theories of cognitive science since the cognitive revolution up until embodied cognition's rise in the past few decades. It is still highly influential. Von Neumann computers are universal processors which at base process binary code and follow deterministic rules, regardless of the high-level veneers of programing languages and user interfaces. At the lowest level there is no ambiguity. These computers are also modular with processes within one delineated processing stream not affecting other processing streams unless specifically instructed to interact. Any universal processor coded with these computational rules, receiving the same input data will have the same output. Essentially, content is device-independent. Major models of general cognitive ability have been built on this premise including SOAR (Laird, Newell, & Rosenbloom, 1987) and ACT-R (Anderson, 1983). Embodied cognition, from its weakest to strongest iteration, suggests cognition is instead inherently device-dependent – content and processes are shaped by the brain, the rest of the body, and the environment. For embodiment theories, cognition is inherently bound to the format of input our bodies are able to interact with and the output our bodies are capable of producing.

2.2.1 Embodiment Through Experience

Many successful cognitive models are based on building up information through experience. Statistical learning capabilities (Saffran, Aslin, & Newport, 1996; Meyer, & Baldwin, 2011), exemplar models of categorization (Goldinger, 1998; Nosofsky, 1988), and cognitive metaphor development (Gibbs, 2006; Lakoff & Johnson, 1999) all suggest knowledge is the product of repeated exposure to information of a particular structure which the cognizer slowly learns to recognize. Our bodies limit our perception – we only perceive information within our *umwelt* - our individual range of perception and interaction with the world (Barrett, 2011). Physical structural relationships such as *optical flow* – the lawful relative motion of elements of a visual scene from the sensor's perspective - lend stability to perception which agents can exploit (Gibson, 1979). Further, repeated experience with the same type of input lends stability to cognitive processing beyond the currently perceptually available information.

Examples from face recognition expertise to color categorization to cognitive metaphor illustrate this stability through repeated exposure. Most humans develop expert abilities at distinguishing faces (Gauthier & Nelson, 2001) by learning what aspects of visual input is most important to distinguish similar faces from each other and what aspects can change while remaining the same face. Across a wide range of faces, humans can recognize basic emotions, i.e. detect an abstract similarity that is not implemented exactly the same on every face, even with only an image of a thin strip of face around the eyes (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001; Baron-Cohen, Wheelwright, & Jolliffe, 1997). In the domain of color, Regier, Kay, and Khetarpal (2007) have shown that color categories across cultures are broadly predicted by optimized divisions of the perceptual signal visible to the human eye. Linguistic and cultural norms influence the categories constrained by the physically guided divisions. Seeing color with the human biological visual system creates some stability and the experience of having some colors treated as equivalent while others are treated as meaningfully different result in the phenomenon of categorical perception (Goldstone & Hendrickson, 2010), when a person can more quickly and accurately distinguish two hues that cross a color boundary than two that are within a color category even if the two hue pairs have the same distance between them in color space. The discrimination advantage is present only for people whose language encodes the color category distinction at a basic level, such as the English distinction between *blue* and *green* or the Russian distinction between *light blue* and *dark blue* that English does not have (Winawer et al., 2007). Finally, cognitive metaphors, while by no means universal, can be viewed as associative primitives within a language and culture. They are built from co-occurrences of concepts. For example, the *future is forward* metaphor captures a physical reality that most directed movement tends to be in the forward direction - the direction faced by the front of the body. The co-occurrence of spatial direction and temporal flow creates a conceptual basis for relating time to space (Lakoff & Johnson, 1980, 1999). The connection of a stable perceptual experience to an abstract concept gives the metaphor of moving towards a goal – a future state – have meaning. Each of these examples demonstrate instances where predictability of sensory input is increased through repeated exposure and that predictability influences perceptual processing. The established knowledge of stable patterns is then used to support abstract conceptual interpretations.

2.2.2 Radical Embodied Cognition

Radical Embodied Cognitive Science (RECS) (Chemero, 2009) is a philosophy that falls within the replacement hypothesis in Shapiro's taxonomy of embodiment theories. It is the most extreme of the embodiment positions believing that not only is the domain of cognition extended beyond the brain but also that the brain is not a particularly privileged part of cognition. The entire agent-environment system, and the unfolding changes within it, are the domain of cognition. This theory, as mentioned above, is explicitly anti-representational.

Many ecological psychology experiments never lead to theories about brain-internal processing mechanisms. The fly ball example introduced earlier is an example of this - the agent only needs to act in a way such that they continue to have a stable percept of the ball in order to catch it. From animal cognition research, there is a similarly non-brain-based explanation for the complex behavior of the Portia genus of jumping, predatory spiders (see Barrett, 2011, ch. 4, for review). These spiders do not have much in the way of neural resources with brains the size of a pinhead. Yet they exhibit incredibly complex behaviors. Observing the hunting behavior of these spiders, it is easy to impose on to them a theory of mind, the ability to conceive of other agents having thoughts and goals. When the spiders are stalking their prey they take a round-a-bout route that might include paths that occlude their target. To accomplish this, don't they have to have a mental representation of the spatial location of their target? Experiments have demonstrated instead that the successful route depends on a simple visual search pattern of tracing unbroken horizontal paths following the simple rules of reversing their visual path tracing back toward the target prey when a break is encountered and continuing to trace a path away from the target if that path is unbroken (Tarsitano & Andrew, 1999). If there is no clear path, the spider selects a salient secondary object to move towards then scan again (Hill, 1979; Tarsitano, 2006). From the observer's perspective, this looks like the spider is carefully planning its actions as it stops and scans back and forth comprehensively. Yet in reality it is always just choosing the next path of motion using this rule. The Portia spider path selection is an example of a simple perception/action loop resulting in very complex, flexible behavior. The Portia spiders also exhibit other simple perception/action loops with complex ramifications while hunting. Portia spiders hunt other spiders. When being furtive, a Portia spider seems to dance across another spider's web in ways that create vibration patterns that mimic natural, ignored stimuli like a breeze on the web allowing them to sneak up on the prey spiders (Wilcox, Jackson, & Gentile, 1996). Alternatively, a Portia spider is able to lure a prey spider to itself by moving on the prey spider's web in a way that is sensed by the prey as the struggles of a caught insect. This signal is accomplished by the Portia spider varying its movements to produce an appropriate vibration pattern. Again, this looks like it could be a planned, knowledge-based behavior. But rather than intentionally mimicking the patterns of other insects, the Portia spider simply varies its motion until it senses a different source of vibration on the web, a vibration indicating the prey spider is on its way to the hunting spider (Jackson & Wilcox, 1990, 1994). If the prey spider stops moving toward the Portia spider, it will again use trial and error to vary its movement patterns until they get the prey spider to continue towards them (Jackson & Wilcox, 1994). Again, this behavior can be accomplished with simple perception/action loops and relies on the environmental cues rather than a cunning mentally represented plan.

RECS builds on these types of findings to suggest that while the brain is clearly part of the agent-environment system, there is no reason to push all of cognition into the brain. From this perspective, the agent uses its capabilities - its sensory apparatuses, its locomotive capacities, the patterns of stability between its body and elements of the environment - to do most of the heavy lifting of everyday cognition.

2.2.3 Embodiment and Representation

Embodiment and representation are distinct but mutually influencing theoretical domains within cognitive science. The extreme stances, such as a rigid interpretation of the Computer Metaphor of Mind and Radical Embodied Cognitive Science, are endorsed by only a few researchers. Most researchers explicitly or implicitly take a position somewhere in between those extremes. The assumptions required of stances on representation and embodiment limit each other. For a researcher who does not take a strong stance on representation, a particular behavior might seem amenable to either being explained by representational means - what Chemero (2009) refers to as mental gymnastics - or by non-representational interactions of the agent-environment system. However, these competing explanations have representational assumptions - each explanation is only compatible with the required representational assumptions of some of the various embodiment hypotheses and directly in conflict with the assumptions of other embodiment hypotheses.

The conceptualization hypothesis relies on representation as a key theoretical construct. In this hypothesis, cognition remains strongly representational, but the representations have embodied properties as core features. Representations are coded using the same brain networks that are used in perception and action. The traditional view is revised from mental representations as amodal conceptual symbols referring to aspects of the external world. The embodied mental representations are inherently modal; they are dependent on the sensory modality of input during formation. In this way, though it changes some aspects of what a mental representation is, the conceptualization hypothesis does not challenge the dominant view that representations are essential components of cognition.

In the constitution hypothesis, representation remains a key component of cognition, but the representations are not defined as exclusively brain-based. The environment contains representations that a cognitive agent can access and use in a functionally similar way to the representations contained within the brain. For this hypothesis, mental representation is not necessarily different from the traditional view of mental representation. The distinguishing aspect is the expansion of what is viewed as a representation – mental and *extended* representations are both part of cognition.

In the replacement hypothesis, representation is not considered a meaningful construct. What is called mental representation by other theories of cognition is a non-discrete component of a physical, causal system. To the extent that a visual stimulus creates a pattern of neural firing, the neural firing is not more important to understanding that cognitive system than the stimulus. It is the whole system - the stimulus, the sensory apparatus perceiving the stimulus, the body moving the sensory apparatus to gather more input from the stimulus and surrounding context, etc. - that makes up the object of study; neural firing patterns exclusive of the rest of the system is too narrow a focus.

These types of embodiment have complex relationships with each other on the subject of representation. The constitution hypothesis is partially compatible with the conceptualization hypothesis. For constitution hypothesis theories, mental representations can be fundamentally influenced by the perceptual means of learning them as must be the case for conceptualization hypothesis theories. The constitution hypothesis goes further, suggesting that many components of meaning lay beyond brain networks in the body and environment. This extension of representation is not required of a conceptualization hypothesis theory. Meanwhile, the constitution hypothesis's connection to the outside world and emphasis on a cognizer's environment seem compatible with the replacement hypothesis. However, there is a fundamental divide over the value of theorizing with mental representations. Both the constitution and replacement hypotheses believe traditional mental representations lack enough explanatory power to be the complete picture of cognition. The constitution hypothesis expands representation to explain more of the world in cognitive terms. The replacement hypothesis rejects representation as meaningful and instead privileges ongoing processes over stable states of a system. Finally, the replacement hypothesis and conceptualization hypothesis completely disagree on representation. Yet, to the extent that there are brain network responses to changes in the agent-environment system, these brain responses would be constituted at least partially of the perception/action networks. The core idea of the structure of the world, as perceived by a cognizer, influencing mental knowledge is shared by both hypotheses. The divide between them is over whether that knowledge continues to contain enough similarity to traditional ideas of representation in order to continue to use that word.

2.3 A Paradigm Shift?

Some theorists have suggested that embodied cognition has the potential to be a unifying theory within psychology (Glenburg, Witt, & Metcalfe, 2013). Others have strongly rejected that suggestion as overreaching (Goldinger, Papesh, Barnhart, Hansen, & Hout, 2016; Mahon, 2014, 2015a, 2015b; Mahon & Caramazza, 2008). Goldinger et al. (2016) suggest the unifying theory claim is a proposal of a paradigm shift on the order of the movement from behaviorism to cognitivism. The unifying theory proposal applies to the general theory of embodiment across all the hypotheses discussed above: the body and environment are essential components of, or at the very least strong influences on, cognition. Changes to the domain of a scientific field, differences in the questions that are meaningful to explore, and the revision of previous terms to mean something new and incompatible with the previous meanings are signatures of paradigm shifts (Kuhn, 1962). In the previous section, it was clear that all branches of embodiment either revise or reject the traditional notion of representation, which according to the listed characteristics could be indicative of a paradigm shift. To be clear, Glenburg et al. (2013) do not make the paradigm shift claim explicitly though Goldinger et al. (2016) explicitly argue against it. Here I will explore the merits for and against viewing embodiment as a paradigm shift.

In order to reject the claim of a paradigm shift for embodied cognition, theorists must defend a position that the phenomena of interest remain fundamentally unchanged within this new perspective. Embodiment theories each argue for some deviation from traditional views of cognition. Whether these proposed differences are enough to be considered a new paradigm remains to be seen. The specific critiques of Mahon and Caramazza (2008), and Goldinger et al. (2016) to the idea of a paradigm shift reject the proposed change in the nature of representations, and reject the expansion of the domain of cognition, respectively.

Within the conceptualization hypothesis version of embodiment, it is often claimed that conceptual information is *grounded* (Barsalou, 1999). That is, knowledge is based on and in sensorimotor processing, and representations are inherently modal. An extensive research program investigating this hypothesis has been undertaken with key researchers including Glenberg and Kaschak (2002); Lakoff & Johnson (1980, 1999); and Hauk, Johnsrude, and Pulvermuller (2004), among many others. Mahon and Caramazza (2008) specifically challenge this grounded representation hypothesis. They put forth the suggestion that if the content of representations can abstract away and no longer depend on sensorimotor areas, the concepts become amodal representations that have little to nothing to do with anything beyond the brain. The critique argues that the keystone finding of sensorimotor areas being active while an individual is thinking of action concepts such as kicking (Hauk et al., 2004) does not conclusively mean the sensorimotor brain areas are integral to conceptual processing; the sensorimotor areas could be activated through spreading activation from amodal representations.

While this criticism may prove useful in restraining potentially overreaching ideas of conceptual processing being exclusively in sensorimotor areas, it does not refute the larger claim of the conceptualization hypothesis that bodily and environmental information is integrated into representations. A concept being abstracted away from an explicitly modal experience is not necessarily amodal - it may contain some influence of the concrete metaphors used to build up to that abstract concept. Concrete, simulatable concepts are learned easier and earlier than abstract concepts (Schwanenflugel & Akin, 1994). The conceptualization hypothesis suggests abstract concepts grow on top of and through appeals to concrete concepts, which are closely based on sensory information. The current focus on visualization technologies for big data to improve understandability illustrates the power of making information concretely perceivable.

Another way to address the Mahon and Caramazza (2008) critique that representation can be amodal given the current empirical literature is to admit that amodal representation could be possible within an embodied account. The language and situated simulation (LASS) theory (Barsalou et al., 2008) claims there are two conceptual systems: one linguistic and one simulation. The linguistic network is shallow and mostly comprised of associations built out of word patterns. The simulation network is the deep conceptual network where grounded concepts are represented. The two networks are interconnected, but associations within the linguistic network, being more narrow in scope, are more quickly activated than associations in the simulation network. This would seem to suggest a separation between words and sensory experience. Abstract concepts could have a larger linguistic component than concrete concepts.

Louwerse (2008, 2010) also uses this framework of a dual network expanding it to suggest that many sensory relations such as typical spatial layouts are encoded into linguistic distributions and therefore would be found in a linguistic network without needing to appeal to a simulation network. A set of experiments demonstrate that images and words both evoke language and simulation factors but the ratio of reliance on one network over the other changes with the stimulus format (Louwerse & Jeuniaux, 2010). Language and simulation are both present but rather than language relying on simulation for its meaning, the language network can encode some information that would seem to be simulation based.

The separation of language networks from sensory networks is an appeal to traditional cognitive science. It would appear that the effect of embodiment for these theories is bounded to particular types of processing and that linguistic representations are of the amodal type. In particular, the influence of simulation on language can be seen as being encoded into the connective structure of linguistic representation networks rather than being essential components of individual representations. This conception would not be considered a radical paradigm shift - most of cognition can still be theorized to work the same way as before embodied cognition was proposed. If embodiment is exclusively of this type then the body and environment are merely an interesting influence on cognition. Cognition theories would continue to be about the same information, attempt to answer the same questions, and use terms with the same meaning.

There is, however, an argument to be made that these results are not indicative of amodal representations: language is experienced in a sensory fashion like all other input into a brain. Language is seen, heard, or, in the case of braille, felt. Finding that language is highly interconnected is unsurprising since linguistic elements are frequently experienced in the context of other linguistic elements. The language network can be viewed as a build up of sensory information like any other visual, auditory, or haptic domain. The rich interconnectedness of language could explain effects such as the Barsalou et al. (2008) finding of faster processing for language network associations. The limited range of possible input for language, comparatively much narrower than the general range of sensory input, creates a more densely interconnected modal network.

Taking a different tack at critiquing embodiment, Goldinger et al. (2016) characterizes embodied cognition to be about body states, environmental states, and in some cases about rejecting representations. The critique also argues that dozens of findings within cognitive psychology do not rely on modal representations, nor are these findings substantially influenced by the type of body containing the cognitive mind. There is a strong online processing bias to this characterization, ignoring embodied theories that focus on the creation of embodied knowledge - the latent relationships not currently part of an active state. Connell & Lynott (2014) suggest that concepts are build up of repeated co-occurrences in active representation resulting in stable connectivity patterns. The history of the cognizer, in a body, within particular environments, guides the development of these concepts. The continued experience of the cognizer continues to influence these concepts that, while stable in a broad sense, are not fixed in the narrow sense. The change of the stable state over time is determined by ongoing embodied experiences.

This idea of history within a body does not fit within the narrow definition of embodiment Goldinger et al. (2016) argue against. In fact, just the idea of an embodied history can challenge most of the identified cognitive findings argued to be unaffected by embodiment. For example, in the cases of frequency effects, exemplar models of categorization, and familiarity of faces facilitating perception, it is the history of perceiving these things that drive the effects and the theoretical explanations. Further, many of the identified findings are in classic cognitive psychology topics studied in research laboratories. The narrow range of stimuli used in a controlled experiment limits the possible environmental influence on results by artificially keeping the environment constant. When elements of a situation are intentionally controlled in order to be statistically ignored, it is not a fair test of whether those elements affect the examined processes. Classic results that could appear to lack an embodied component are judgment and decision-making biases such as the anchor effect (Tversky & Kahneman, 1974), and the hindsight bias (Fischhoff, 1975). In both of theses biases, judgments are affected by the information taken in prior to the judgment - information context effects. A similar effect, estimating distance and incline grade (Proffit, 2006) from the bottom of a hill is affected by the current energy level of the person making the judgment. The judgment is influenced by information that is stored in the body, not represented in the brain. The anchor and hindsight bias effects may show embodied influences if tested in a context in which they would be identifiable.

Even if embodiment can be considered a paradigm shift, taking an embodied view does not mean all results will change and be directly affected by the overarching idea that the body bounds perceptual input. Familiarity and frequency are determined by the long-term environment a cognizer is situated within. The relative frequencies of particular bands of x-ray waves in their raw form will never be familiar and never have consequences for categorization among human perceivers because they are beyond routine human sensory capacities. An organism is only able to gather input from within its umwelt - its range of sensory experience; other information is not directly perceivable and needs to be transformed in order to be perceived, such as taking x-rays and printing them to have contrasts within the human visual spectrum. Goldinger et al. (2016) suggest that if embodied cognition is the observation that we have bodies, that is profoundly uninteresting. I would argue instead that it would be profoundly uninteresting if the study of cognition is the study of human behavior exclusively within artificially controlled settings. Using an overly narrow definition of embodied cognition to argue against the general theory of embodiment is a rhetorical mistake.

Further, one signature of a paradigm shift is the changed meaning of scientific terms. Goldinger et al. (2016) reject the replacement hypothesis out of hand as incoherent. Representation is such an integral component of traditional cognitive theories. A cognitive theory without representation seems impossible while entrenched in the assumptions of that paradigm. In the conceptualization hypothesis, the concept of representation is stretched a bit to include sensory information. That stretch does not necessitate the reconceptualization of the field, a true paradigm shift. The constitution hypothesis also stretches representation, including more things as representational. This is a substantial difference in what a cognitive representation is when compared to traditional cognitive science. Whether this difference is enough to be consider a true change in the term's meaning, or if it is an adjustment, is debatable. The replacement hypothesis, however, in rejecting the term representation is asserting a fundamental change in what brain network connections mean. What traditional cognitive science calls mental representation and the role they fill in cognitive theories is rejected. Network connections and network activations are not quasi-symbolic mental objects. If the word 'representation' were applied to the network attributes within the replacement hypothesis, representation would not mean the same thing as it does in traditional theories.

Similarly, within embodiment the meaning of the term cognition itself changes to encompass more processes than was traditionally considered the domain of cognition. Non-embodied cognitive science remains entrenched in the view that perception is cognitively impenetrable (Firestone & Scholl, 2015; Pylyshyn, 1999). This means that there is a distinction between perceptual mechanisms and cognitive mechanisms within the brain. For all versions of embodiment, cognition is not exclusively the processing that occurs after perception and before action. It is more expansive with substantial interactions across these traditional divides. It is unclear how to distinguish where perception, cognition, and action transition with each other. In fact, finding points of transitioning between these categories of processing is no longer a meaningful and interesting question.

In embodied cognitive science, action-perception loops can be seen as the programming language of the brain. Through action-perception loops, brains develop processing structures that support functional, meaningful future action. Processing structures do not need to be symbolic in that they do not have to stand for something. They can simply be the built up response to stimulation that unfolds to produce meaningful action without there being a spot in the brain one can point to as the 'representation' of that stimulus. Within the replacement hypothesis, the physical networks of the brain are viewed as causally related structures corresponding to brain-external aspects of the agent-environment system. While it might be possible to point to something representation-like, the delimitation of the brain processing from all of the other situated environmental processes is not meaningful. This version of embodiment acknowledges that brains are not universal processors. The mind is not simply software that can be implemented on any sufficiently powerful hardware. The hardware of cognition – the brain within the body within an environment – affects the contents of cognition. Because this conception of processing is radically different from traditional symbol-and-rule processing, or even the more modern connectionist distributed symbols with complex interaction networks, researchers entrenched in the traditional paradigm would view the repurposing of the processes as incoherent. In a way, the rhetorical argument against the replacement hypothesis in the Goldinger et al. (2016) critique is evidence for a change of meaning substantial enough to be considered a paradigm shift.

Whether embodiment could be a paradigm shift is an open question. Embodiment remains a large range of hypotheses with some having representational commitments that are a narrow tweak from traditional cognitive science to others having broad fundamental changes such as in the amply named Radical Embodied Cognitive Science. A theoretical stance on cognition which includes the agent-environment system would ask vastly different questions from those cognitive scientists asked in the past. The full context becomes a key component of all theories rather than part of a specialized niche of research. It is safe to say that at least the replacement hypothesis, and possibly the constituent hypothesis, would represent a paradigm shift. Whether these hypotheses will win out and become standard within cognitive science over the non-embodied theories remains to be seen.

Chapter 3

Representational Shifts Made Visible: Movement Away from the Prototype in Memory for Hue[3]

3.1 Introduction

Memory, reasoning, and categorization have traditionally been distinguished as separate topics and separate areas of research (Hayes et al., 2014; Heit & Hayes, 2005; Heit et al., 2012). It could be argued that categorization is either an automatic process as in categorical perception, where conscious reasoning is not recruited and any effect of categories on perception would appear to be due to the activation of categorical memory, or alternatively, categorization is an explicit process as in categorical decision tasks where a more deliberative process of reasoning may be at work. But the dichotomies of implicit vs explicit and memory-based vs reasoning-based categorization are too extreme; instead, a continuum is likely to exist. We suggest that that these cognitive activities are intimately intertwined. As Churchland (1981) pointed out, the terms we have from folk psychology, the ways we culturally divided cognitive processes prior to having scientific evidence to inform those divisions, are not necessarily sensible. As research advances in psychological and neurological understandings of cognitive processes, these traditional terms and divisions need to be broken down. The memory, reasoning, and categorization distinctions are losing their usefulness as separate constructs due to the likelihood of common underlying mechanisms.

In this paper, we will be looking at a task that involves both memory and categorization. Experimental participants either label hues with basic color categories or make preference judgments about the hues. Then, memory for these hues is tested immediately. Participants have memory of categories and, through categorization, bring that memory to bear on newly formed encodings. In the way categorization and memory are often discussed, categorization is the act of applying knowledge while memory is the substance of that knowledge. Yet using memory edits memory itself, as has been shown with memory reconsolidation (Nadel et al., 2012) and retrieval induced forgetting (Anderson et al., 1994). Memory and categorization cannot be treated as fully distinct cognitive topics but are interdependent.

The distinction between perception and memory is also a vague and possibly false distinction. Perception is the transduction of light, sound

[3] The official citation that should be used in referencing this material is Kelly, L. J. & Heit, E. (2014). Representational shifts made visible: Movement away from the prototype in memory for hue. *Frontiers in Psychology, 5*.

waves, chemicals, pressure, and heat into electrical signals in the nervous system. Memory refers to the storage of that information. Milliseconds after a stimulus has been experienced, researchers consider it remembered in iconic memory, some of which passes on to working memory and possibly to long-term memory. There has been a debate about how far top-down conceptual knowledge can impact perception with some researchers arguing that perception is cognitively impenetrable (Pylyshyn, 1999) and others arguing that cognitive expectations affect perception very early in processing (Churchland, 1988; Hsieh et al., 2010).

One of the main phenomena of interest in the cognitive penetrability debate is categorical perception, where categorical knowledge affects how people perceive the surrounding world. Categorical perception has been examined in many domains including phoneme perception (Liberman et al., 1957), faces (Levin & Beale, 2000), and color (Winawer et al., 2007), (see Goldstone & Hendrickson, 2010 for a full review). In categorical perception, there is no deliberative reasoning – categorization is implicit and automatic, having an effect without people needing to actively decide on a category. Here, categorization appears to be based on implicit memory of frequent categorizations.

Categorical perception has been explained diversely: as a pull towards the prototype (Luypan, 2008), a truncation at the boundaries of a category (Huttenlocher et al., 1991), or an expanding of perceptual space (Goldstone, 1994, 1998). None of these accounts of categories on perception would explain the novel result we present here: With rapid presentation and test of hues there is an atypical bias – a push away from the prototype, a pull towards the boundary, or a seemingly incompatible change of perceptual space. While this result is novel for hue memory, a similar effect has been observed in immediate recognition of exaggerated faces (Rhodes et al., 1987). During perception, people appear to bring categories to bear on the content of perception but the influence is not uniformly one of attraction toward the prototype.

Our own investigation was spurred by the argument that labels affect the memory of perception when the labels coincide with perception (Lupyan, 2008). Specifically, for an effect that was metaphorically referred to as a representational shift, it was claimed that labels cause memory traces to be prototypically shifted from the raw percept by exerting a top-down influence of the labeled category on the perceived item. The label activates the category prototype, which interacts in real time with the bottom up perception resulting in a mixed encoded memory trace. Specifically, these experiments looked at whether there was an advantage to remembering objects that were labeled or judged in terms of preference (liking). The participants either labeled object categories including chairs, lamps, and tables (two categories per experiment) or made a like/dislike preference judgment in alternating blocks during study. Participants only saw the objects for 300ms and had 700ms to respond to discourage labeling in the preference judgment trials. After all study trials, participants were then tested on their memory for the items using the original objects as well as a matched lure for each original item in a new/old recognition task. Participants less accurately remembered previously seen items if they had been categorically labeled, which was taken as evidence that the representation of the labeled objects was shifted – it no longer matched up to the originally perceived item. Other researchers (Blanco & Gureckis, 2013; Richler et al., 2011) have taken issue with this interpretation in terms of representational shifts, instead suggesting that perceived items are remembered better because preference judgments require a greater depth of processing than category labeling. They introduced non-labeling conditions such as chair orientation (Blanco & Gureckis, 2013) and screen position (Richler et al., 2011) that only require superficial processing of the objects. These conditions performed similarly to the category labeling condition introduced by Lupyan (2008). To these researchers, the strength of the memory accounts for the differences in recognition memory.

The controversial claim from Lupyan (2008) that there are prototypical representational shifts has not been demonstrated directly. Only a decrease in accurate recognition of previously seen items has been shown, which could mean a shift toward the prototype, away from the prototype, or simple forgetting without a directional change in the memory trace. Hence, previous research on representational shifts has not provided clear evidence that representations have shifted, much less in what direction they have shifted. To get at this question we will present a paradigm that is a conceptual replication of Lupyan (2008) using the same judgment conditions at target presentation, category labeling and preference judgments, and using a similar memory test, same/different judgments rather than new/old recognition judgments. The main differences are in stimuli and timing. We present the targets as well as four matched lures varying systematically in category typicality and distance from the target. These stimuli will allow us to quantify the direction and magnitude of any representational shifts that occur. If a shift is in the typical direction as predicted by Lupyan (2008), the new array of test stimuli will allow it to be seen.

Previous work on representational shifts has examined memory for objects such as lamps and chairs. In our own work, we focus on color space, which is more quantifiable and better-defined than object space. Color is a continuous uniform physical space made up of different wavelengths of light. Color is also a rich psychological space that is divided into superordinate, basic, and subordinate categories. There are focal or prototypical colors within categories as well as boundaries where one category meets the next that are shared amongst speakers of the same language, and to some extent across languages (Berlin & Kay, 1969; Regier & Kay, 2009). As such, color space is a fertile testing ground for examining how categorical knowledge distorts basic perception.

Taking account of the psychological landscape of the color domain and people’s ability to detect fine alterations from one color stimulus to the next, we were able to directly test the color that has been encoded through a recognition test, and how different, if at all different, the encoded color is from the originally presented color. By moving from object space to hue space, we constrain the potential directions of memory shifts toward or away from the prototype. By testing memory of the target as well as four matched lures differing in distance and direction in hue space relative to the prototype, we have the opportunity to measure the sensitivity (d') of hue memory at different locations relative to the target. Sensitivity serves as a measure of confusability and strength of confidence in having seen something at the point in hue space. Where d' is high, people can reject lures that they have not seen. Where d' is low, this means that lure items nonetheless seem relatively familiar, as if there is a false memory representation at that point in hue space. Moreover, if d' is lower in one direction, relative to the prototype, compared to the other direction, this implies that the representation in memory has shifted along the hue dimension. Our paradigm allows us to see the shift as well as quantify its direction and strength.

Additionally, the representational shift hypothesis focuses on encoding. However, the paradigm used in the original paper (Lupyan, 2008) as well as the versions of the paradigm used in subsequent work (Blanco & Gureckis, 2013; Richler et al., 2011, 2013) have an extended study phase presenting all items twice prior to a test phase of all items resulting in a delay of minutes between presentation and test. This format does not isolate effects down to the time of encoding. Our paradigm focuses on immediate memory to more closely address encoding. We use a same/different judgment as the memory test either 500ms after target presentation (Experiments 1a and 1b) or 5000ms after target presentation (Experiment 2a and 2b). This prevents interference of other hues on the representation of the key item between study and test.

We now present four experiments, two main experiments and two direct replications. Experiment 1a was designed to test the memory for a color soon after encoding. The delay between original presentation and the same/different judgment was 500ms. We found an atypical shift—an unexpected finding based on previous research which had suggested that the shift would be towards the prototype—with no difference between judgment conditions. In Experiment 2a, the delay was increased to 5000ms to test if the predicted prototypical shift could be observed at a longer delay and if there was an effect of judgment condition that developed over time. The atypical shift and lack of judgment condition effect were reproduced. Due to some participants being excluded from Experiments 1a and 2a as well as the unexpected direction of the representational shift, we conducted direct replications of both experiments with higher power, in Experiments 1b and 2b.

3.2 Experiment 1a

In a conceptual replication of Lupyan (2008), we presented participants with hues to be judged either by category or by preference. In a departure from the previous paradigm that had separate study and test phases, test occurred immediately after study within a trial. Given that the representational shift hypothesis is one of shift at encoding, shifts should be immediately detectable. Additionally, rather than having one matched lure for each studied item, there were four lures spanning both potential directions of movement relative to the prototype and two distances in hue space. Using sensitivity (d') as the dependent measure, we determined whether memory shifted at all, if it shifted towards the prototype or away from it, and the approximate distance of the shift in hue space. In particular, lower d' values indicate a higher false-alarm rate to lures. So, for example, if representations shift towards the prototype, there will be greater likelihood of false-alarming to typical lures compared to atypical lures, and d' will be lower for typical items than for atypical items.

3.2.1 Method

3.2.1.1 Participants

Thirty-six students at the University of California, Merced participated in these experiments for course credit. All participants reported normal vision and normal color vision. Their color vision was tested using the CITY colorblindness test (City University, 2002) following the main experiment. The research was approved by the University of California, Merced Institutional Review Board and verbal consent was obtained from each participant.

3.2.1.2 Materials

The color stimuli were calculated in CIE L*CH color space then translated to CIE L*ab color space. The stimuli were from two color categories, red and green. Focal colors, treated as the category prototypes, were obtained from Sturges and Whitfield (1995). Saturation and brightness were held constant at the focal saturation and focal brightness. Within these color categories, four target colors were selected for a total of eight target colors across the two categories. All target hues were of similar typicality relative to the prototypes though explicit typicality measures were not collected. The targets were neither extremely typical nor atypical of their color category. From each of the target colors, four variants were created, two closer to the prototype and two further away from the prototype. These variants served as the recognition test lures. The hue distance between each hue in the set of 5 test hues, the target and four lures, was equal. The hue distances were normalized for the different color spaces with green encompassing a larger number of degrees than red. All variants within a set did not cross the prototype or the color category boundaries. The calculated colors can be found in Chapter 4, Section 7.2.

Dell Ultrasharp U2410 monitors were used to display the stimuli and the color calibration profiles were created using a X-rite i1 Display Pro color calibrator. The stimuli were created using Adobe Photoshop to convert the calculated colors to a RGB device specific color profile for each monitor, resulting in uniform presentation across the three monitors.

Using a photometer, the experimental cubicles were found to have similar intensities of light from the overhead fixtures.

3.2.1.3 Procedure

There were two target judgment conditions. Participants chose between the basic color categories, green and red, for the categorical judgment and between like and dislike for the preference judgment. The categorical judgment response keys were counter-balanced across participants whereas the like/dislike response keys were in left to right order as it is a natural mapping. The second judgment of each trial was a same/different judgment. The participant was to judge whether the second hue presented during the trial matched the first hue that had elicited the category or preference judgment.

Each trial consisted of a fixation cross (1500ms), the target hue (300ms), a question mark eliciting a button push judgment (up to 700ms), a blank screen (500ms), and a response screen with a test hue also eliciting a button push judgment (up to 4000ms). The participant's response immediately ended the response-eliciting screens. The trials were portioned into blocks of 80 trials consisting of all 8 target colors being paired with each of their 5 test hues (the original hue and the 4 lures) for 2 trials. Each block had one type of judgment (categorical or preference) being the response to the question mark. There were 4 blocks, alternating between the judgment types. The order of the blocks was counter-balanced across participants.

Prior to the main trials, participants were trained on 6 yellow and purple stimuli trials, then completed 2 short blocks of 10 trials, one block of category judgments and one block of preference judgments to allow participants to get into the rhythm of responding quickly before the key trials began. These short blocks contained the red and green stimuli and were not indicated to be practice trials to the participants.

3.2.2 Results and Discussion

We excluded 12 participants, 8 for low color naming accuracy (<80% correct) in spite of color vision screening, as well as 4 for failing to follow instructions. Failing to follow instructions in this and subsequent experiments included a very high rate of 'like' judgments >90%, a low response rate at either the judgment or test portion of a trial (<80%), or always responding with 'same' at test. Including these participants does not change the pattern of results. Here, we report results based on the 24 remaining participants. In this and subsequent experiments, individual trials were excluded if categorization at study was incorrect or participants did not respond at both the study and test portions of a trial.

The analyses relied on the d' measure of sensitivity used in signal detection theory (Stanislaw & Todorov, 1999). The d' measure has been used in recent studies of representational shift (Blanco & Gureckis, 2013; Richler, Gauthier, & Palmeri, 2011; Richler, Palmeri & Gauthier, 2013) though false alarm rates alone were used in the original paper (Lupyan, 2008). Compared to analyses based on raw scores such as false alarm rates, d' not only takes account of variations in hit rate but has the advantage of being a better match for the underlying Gaussian nature of recognition data (see Macmillan & Creelman, 2004, for a general overview, and Heit & Rotello, 2014, for a more recent discussion). For the analyses, we calculated two overall hit rates per participant one for each condition and used these along with the four lure false alarm rates per condition to calculate d' values. By using d' rather than raw false alarm rates, we are controlling for the general response rate of an individual in a condition in addition to calculating how well they can differentially respond to the target vs. the lures.

In this case, we had one set of test items that were the same as the originally presented item. Same judgments on these items were considered hits and different judgments were considered misses. We also had 4 sets of items that were different from the original hue varying in hue space distance (1 step or 2 steps) and in direction of typicality (more typical or more atypical of the color category). Same judgments in response to these items were false alarms and different judgments were correct rejections. We calculated d' by subtracting the z-score of the proportion of false alarms from the z-score of the proportion of hits. In the case of proportions of 0 and 1, z-scores cannot be calculated due to the normal curve expanding to infinity at its tails. We used the standard correction of including or excluding half a hit or half a false alarm where appropriate (Snodgrass & Corwin, 1988). The hit and false alarm rates for all experiments are reported in Table 3.1. The d' measure was calculated for each of the four levels of the test hue variations and for each of the two judgment conditions by subject.

A significantly lower level of sensitivity, namely a lower d', was taken to be evidence of the direction of a shift. Lower d' corresponds to more false alarms or more non-targets confused to be the same as the target. So, for example, if there is lower d' as a result of more false alarms for prototypical items, this suggests a prototypical shift – the memory traces are treated as more similar to the more typical test items than the less typical items. Likewise, lower d' for atypical items suggests an atypical shift. No difference, or a symmetrical sensitivity, would imply that memory does not shift relative to the category typicality gradient. The representational shift hypothesis (Lupyan, 2008) suggests that there should be an interaction of typicality and condition with a lower d' for the more typical lures than the atypical lures, only in the category labeling condition. There should be no typicality effect, or at least a smaller effect, for the preference condition. The depth of processing account (Blanco & Gureckis, 2013; Richler et al., 2011) predicts a main effect of condition with no typicality effect; the sensitivities should be symmetrical. The key prediction of this account is more accurate memory, a higher d', for items in the preference condition which is suggested to be more deeply processed than the items in the category labeling condition.

Table 3.1

Hit and False Alarm Rates by Experiment and Judgment Conditions

		Hue Type				
Experiment	Condition	-2	-1	Target	1	2
1a	Category	0.638	0.770	0.805	0.730	0.518
	Preference	0.693	0.770	0.811	0.698	0.573
1b	Category	0.660	0.791	0.816	0.750	0.589
	Preference	0.636	0.801	0.827	0.731	0.615
2a	Category	0.646	0.782	0.743	0.655	0.490
	Preference	0.600	0.739	0.715	0.624	0.491
2b	Category	0.574	0.618	0.729	0.597	0.511
	Preference	0.576	0.651	0.662	0.595	0.482

Note. All numbers are proportions of hits to misses or false alarms to correct rejections. The target column represents the hit rate while the other columns represent the matched lures. Sign denotes towards (+) or away (-) from the prototype and number (1, 2) denotes steps distant from the original hue.

To test whether there was a difference in d' by condition (color vs preference judgment), distance from the original hue (1 step vs. 2 steps in hue space) or by direction of typicality (typical vs. atypical), we ran a 2x2x2 ANOVA. The results can be seen in Figure 3.1. There was no effect of condition, with labeling the category or making a preference judgment not differentially affecting hue sensitivity. There was a main effect of distance, $F(1,23)=29.59$, $p<0.001$, $\eta^2=0.184$. The distance of two units had a d' mean of 0.672 while the d' mean of the distance of one unit was 0.259, indicating that there was less sensitivity to a hue change when the test hue was closer to the original hue, as one would expect. This finding indicates that the hues that are less different in color space are less detectable. Therefore any shift that has taken place with the color hues is subtle and within a few degrees of hue space.

The key finding was a main effect of typicality, $F(1,23)=12.92$, $p<0.01$, $\eta^2=0.100$. The d' mean of more typical test hues was 0.612 and the d' mean of hues less typical of the color category was 0.318, indicating that participants were less sensitive to changes in hue if the hue was atypical of the color category. In other words, participants were more likely to false-alarm to atypical test items than to typical test items. Based on prior theoretical work, the prediction was actually the opposite, that there would be less sensitivity, and therefore a representational shift, in the typical direction.

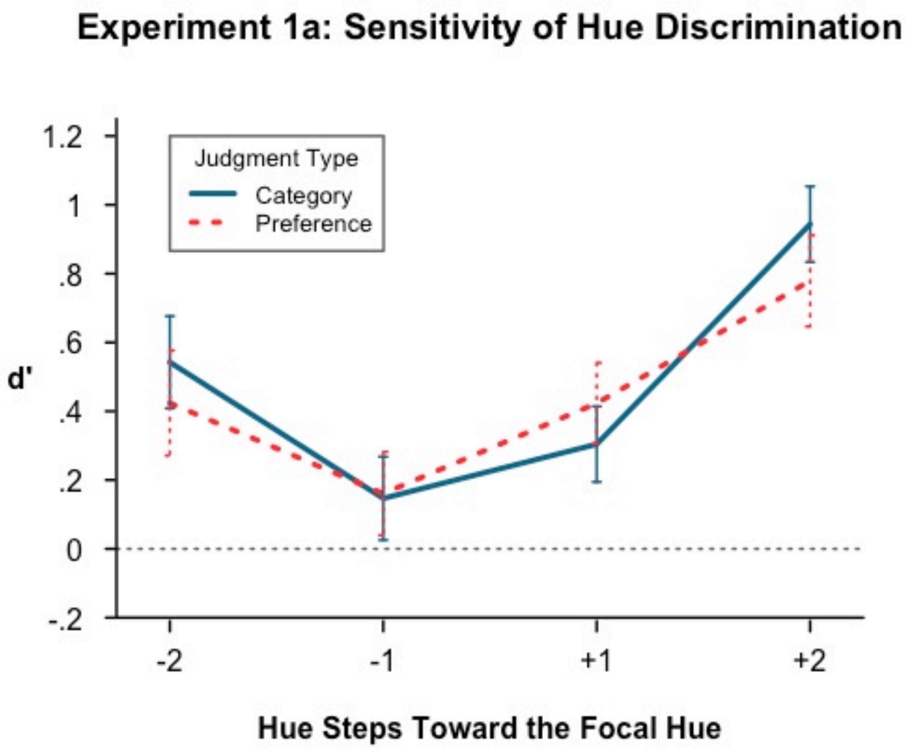

Figure 3.1. Sensitivity of hue discrimination in Experiment 1a. Lower d' values indicate more confusion of hues at a particular level of the hue factors, here denoted with sign for towards (+) or away (-) from the prototype and with number (1, 2) for steps distant from the original hue. Here, d' is lower in the atypical (-) direction than in the prototypical. It is also lower at 1 step removed from the originally presented hue than at 2 steps. There is no difference by condition. Error bars represent the standard error of the means.

Additionally, there were two marginal components of the ANOVA, an interaction of condition and distance, $F(1,23)=3.11$, $p=0.091$, and an interaction of distance and typicality, $F(1,23)=3.11$, $p=0.112$. Due to the unexpected result, the marginal findings, as well as the number of participants excluded resulting in a small final sample size, we view conducting a direct replication as important to having confidence in our results (see Cesario, 2014 for discussion of the importance of direct replication).

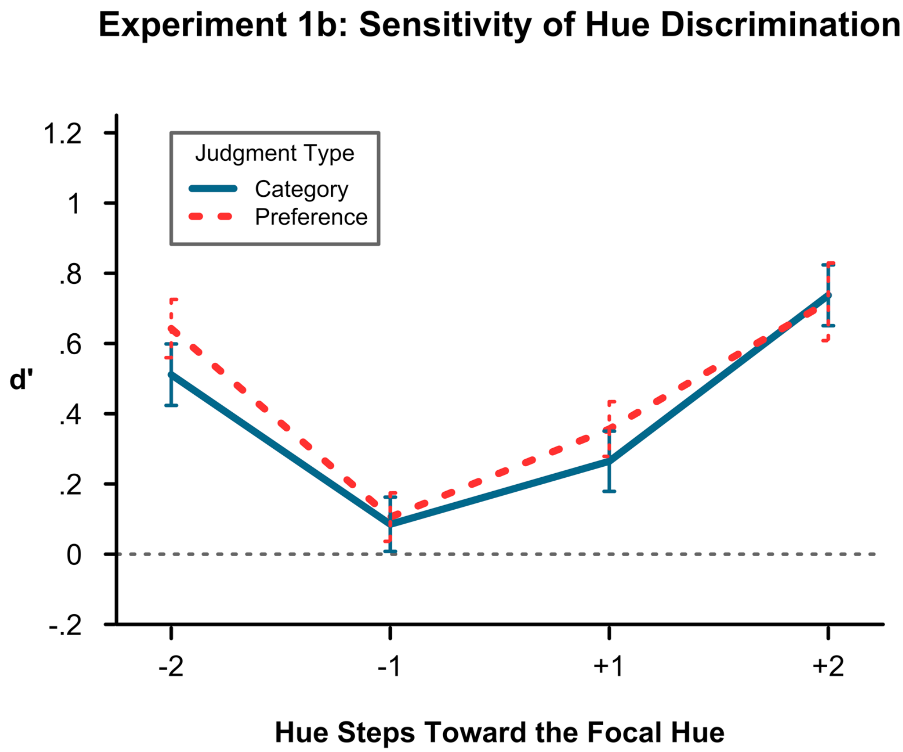

Figure 3.2. Sensitivity of hue discrimination in Experiment 1b. Lower d' values indicate more confusion of hues at a particular level of the hue factors, here denoted with sign for towards (+) or away (-) from the prototype and with number (1, 2) for steps distant from the original hue. Here, d' is again lower in the atypical (-) direction than in the prototypical direction. It is also lower at 1 step removed from the originally presented hue than at 2 steps. There is no difference by condition. Error bars represent the standard error of the means.

3.3 Experiment 1b

3.3.1 Method

Sixty-seven students participated using the same criteria as in Experiment 1a. All materials and procedures were the same.

3.3.2 Results and Discussion

We excluded 22 participants, 10 for low color naming accuracy (<80% correct) in spite of color vision screening, as well as 12 for failing to follow instructions. Including these participants does not change the pattern of results. Here we report findings based on the 45 remaining participants.

The results are shown in Figure 3.2. Conducting the same 2x2x2 ANOVA on the d' scores, there was again no effect of condition. We replicated the significant main effect of distance, $F(1,44)=83.58$, $p<0.001$, $\eta^2=0.242$, indicating that again it was easier to distinguish items further from the original with the d' mean of 2 units being further from zero at 0.652 than the mean of the 1 unit hue distance items at 0.203. We also replicated our typicality main effect, $F(1,44)=10.39$, $p<0.01$, $\eta^2=0.048$, with the atypical direction mean of 0.336 and the typical direction mean of 0.519. Again, participants were less sensitive to changes in hue if the change was away from the category prototype. None of the interactions were significant including the previously marginal results.

With this replication we can have more confidence in concluding that the representational shift is occurring in the atypical direction.

3.4 Experiment 2a

The representational shifts at a half second delay between presentation and test in Experiments 1a and 1b were in the atypical direction. Lupyan (2008) had argued that labeling should have an effect of increasing the typicality of a representation at encoding. Here, we increased the delay between presentation and test to 5 seconds to see if a labeling effect or a reversal in the direction of the shift emerged with more processing time.

3.4.1 Method

Forty students were recruited as in the other experiments. The materials and procedure were the same as in Experiments 1a and 1b with the exception of increasing the delay of the blank screen between the original hue presentation and the test hue presentation from 500ms to 5000ms.

3.4.2 Results and Discussion

We excluded 21 participants for low color naming accuracy (<80% correct) in spite of color vision screening. Including these participants does not change the pattern of results. Here, we report results based on the 19 remaining participants.

We conducted a 2 (color vs. preference judgment) x 2 (1 hue step vs. 2 hue steps) x 2 (typical vs. atypical direction) ANOVA on d' as in the previous experiments. Figure 3.3 shows the d' means and error at each level of the ANOVA. There was no effect of condition. We found a main effect of distance ($F(1,18)=54.18$, $p<0.001$, $\eta^2=0.297$) with the 2 units of hue distance being more detectable (mean = 0.513) than the 1 hue step (mean = 0.057). We also again found a main effect of typicality ($F(1,18)=21.99$, $p<0.001$, $\eta^2=0.204$) with more atypical hues being less detectable (mean = 0.1) than more typical hues (mean =0.47). No interactions were significant.

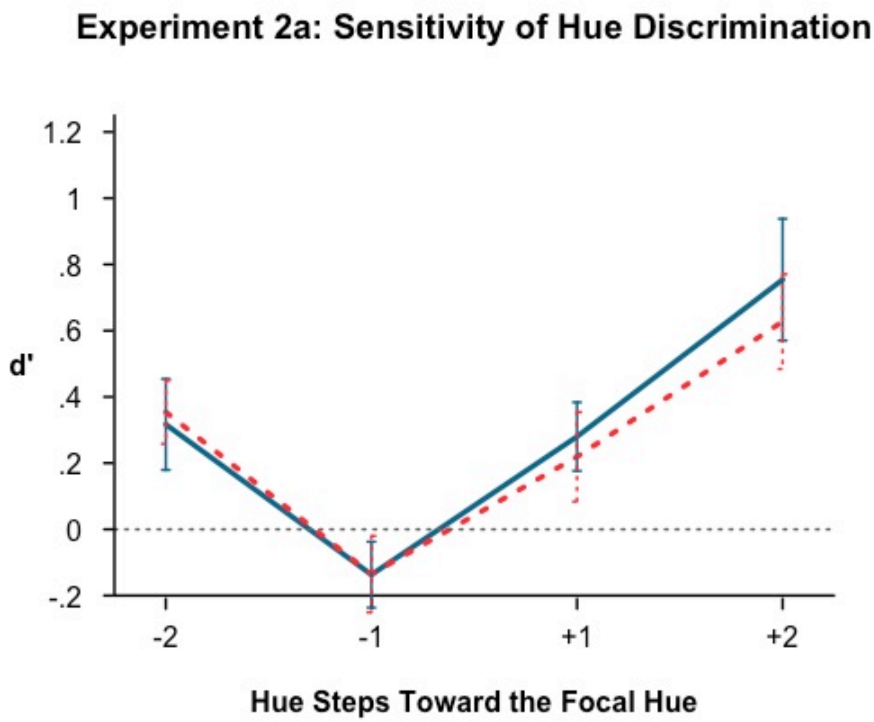

Figure 3.3. Sensitivity of hue discrimination in Experiment 2a. Lower d' values indicate more confusion of hues at a particular level of the hue factors, here denoted with sign for towards (+) or away (-) from the prototype and with number (1, 2) for steps distant from the original hue. Here, d' is again lower in the atypical (-) direction than in the prototypical direction. It is also lower at 1 step removed from the originally presented hue than at 2 steps. There is no difference by condition. Error bars represent the standard error of the means.

As in Experiments 1a and 1b, there was no effect of condition and a significantly lower sensitivity in the atypical direction, pointing again to a shift away from the prototype. However, there was again a relatively high number of exclusions resulting in a low final sample size. We again conducted a direct replication.

3.5 Experiment 2b

3.5.1 Method

We recruited 62 participants using the same criteria as the previous experiments. All materials and procedures were the same as in Experiment 2a.

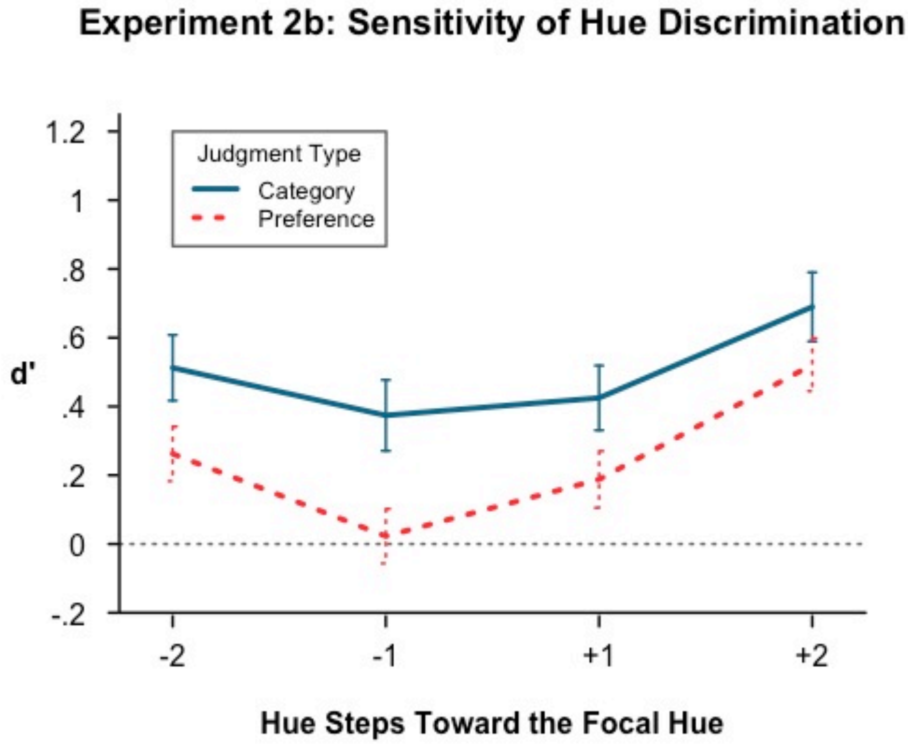

Figure 3.4. Sensitivity of hue discrimination in Experiment 2b. Lower d' values indicate more confusion of hues at a particular level of the hue factors, here denoted with sign for towards (+) or away (-) from the prototype and with number (1, 2) for steps distant from the original hue. Here, d' is again lower in the atypical (-) direction than in the prototypical. It is also lower at 1 step removed from the originally presented hue than at 2 steps. There is a difference by condition with less sensitivity in hues that were preference-judged over hues that were category-labeled. Error bars represent the standard error of the mean.

3.5.2 Results and Discussion

We excluded 14 participants, 7 for low color naming accuracy (<80% correct) in spite of color vision screening and an additional 7 for failing to follow instructions. Including these participants does not change the pattern of results. Here, we report results based on the 46 remaining participants.

We again conducted the same 2x2x2 ANOVA on the d' measurements (Figure 3.4). There was a main effect of judgment condition in this experiment unlike the 3 others, $F(1,45)=6.26$, $p<0.05$, $\eta^2=0.067$. Preference-judged items (mean=0.249) were less detectably different from the original than color judged items (mean=0.500). In other words, there were more false alarms, and more shifting, for preference judgments than for labeling judgments. There was also a main effect of distance ($F(1,45)=30.60$, $p<0.001$, $\eta^2=0.103$), replicating the finding that items 2 hue steps distant from the original (mean = 0.496) are more detectable than items 1 step distant (mean = 0.252). We also replicated our previous main effect of typicality ($F(1,45)=7.48$, $p<.01$, $\eta^2=0.043$) with more atypical items being less detectable (mean = 0.293) than more typical items (mean = 0.456). Again, no interactions were significant.

Whereas d' is a direct measure of discrimination, for completeness we conducted post-hoc analyses of the raw false alarm rates using the same 2x2x2 ANOVA. (See also Table 3.1.) In all of the previous experiments reported here, the pattern of results was consistent with the ANOVA conducted on d' scores. In this experiment, while the false alarm rates of the foils continued to show the typicality effect, $F(1,45)=8.688$, $p<0.005$, and the distance effect, $F(1,45)=31.1$, $p<.0001$, the main effect of condition was not significant when false alarm rate was the dependent variable, $F(1,45)=0.006$, $p=0.939$. The main effect of condition in the d' analysis was driven by taking into account the participants' high labeling condition hit rates.

Overall, we replicated the direction of the shift. That is, extending the time between study and test from half a second to five seconds did not change the atypical nature of the representational shift. Here we observed a difference between conditions for the first time. Category-labeled hues were more distinguishable from the original than those that were preference-judged. We hesitate to draw too strong a conclusion from this difference since it was only found in one of the four experiments and the false alarm analysis implied that the condition effect depends on the hit rate rather than the false alarm rate. Interestingly however, with only a main effect and no significant interaction, the category labeling increased sensitivity to hue differences overall, rather than in a particular direction towards or away from the prototype.

3.6 General Discussion

In our experiments, we found less sensitivity to differences between a studied hue and an unstudied test hue that is less typical of the category. In other words, participants were more likely to false-alarm to atypical items than to typical items. We take this as evidence that there are representational shifts and that they are away from the prototype. Additionally, there may be a judgment condition effect that emerges over time, with category labeled hues being more easily detected as different from the original hue. There is no interaction of the condition and the typicality direction indicating that while category labeled hues might be more detectable, it is not due to shifting. Instead, we can speculate that labeling a color allows participants to reduce bias in either direction equally.

3.6.1 Representational Shifts, Depth of Processing, or Transfer Appropriate Processing?

The representational shift hypothesis (Lupyan, 2008) predicted that (1) there are representational shifts, (2) the shifts happen at encoding, (3) shifts are in the prototypical direction, and (4) category-labeled items are more strongly shifted. Our paradigm was able to show that there are representational shifts and that they occur very quickly, possibly at encoding. However, we found atypical shifts instead of the predicted prototypical shifts. Additionally, we found shifts for both category labeling judgments and preference judgments, indeed in Experiment 2b with stronger shifts in the preference condition. Given these results, we question whether the representational shift hypothesis as detailed by Lupyan is the appropriate explanation here.

Depth of processing (Blanco & Gureckis, 2013; Craik & Tulving, 1975; Richler et al., 2011) does not fit the present results either. If preference judgments require greater depth of processing, and depth of processing leads to better memory, then preference judgments should lead to more sensitivity to hue changes. In fact, we found that preference judged hues led to either indistinguishable sensitivity or less sensitivity than category-labeled hues. Therefore, the depth of processing account of the representational shifts is not satisfactory either.

Instead, we point to transfer appropriate processing (Morris et al., 1977) as a potential framework in which to understand the results. In transfer appropriate processing, relevant details to the task at hand are processed with more depth than details that are less appropriate at the time of encoding. Perhaps preference judgments have an inherently greater depth of processing compared to basic categorization, but the content of that depth is not necessarily what is needed for greater sensitivity in the present task. Directing processing into a comparison of the hue against preferences and making a valence judgment may distract from the encoding of the exact hue while color labeling concentrates processing on the appropriate aspect of the hues for greater sensitivity. Rather than greater raw processing, the right kind of processing leads to more exact memory.

3.6.2 Atypical Shifts

Counterintuitively, the representational shifts at a rapid test pace were in the atypical direction. Previous research on categorical knowledge effects on memory mostly suggests that if memory is altered systematically from the original percepts it should be in a prototypical direction (e.g, Heit, 1997). Categories serve to generalize our knowledge and to highlight similarity among distinct exemplars. What purpose could be served by atypical shifts? While we will not claim to have a final answer to this question, we speculate that it is related to perceptual expertise processes.

A domain where a similar atypical representational shift has been found is in recognition of faces. This has been called a distinctiveness effect (Rhodes et al., 1987). Participants were faster to recognize exaggerated faces over the original facial proportions that were in turn recognized faster than more generic versions of the faces. The authors argued the most distinctive features of a face are what are encoded into memory with the more generic portions not encoded as strongly. Gist memory (Reyna & Brainerd, 1995) can be used to fill in the representation. When a person then goes to use the encoding to recognize a face, the exaggerated face matches the distinctive features better than the true face.

Hue is much less complex than faces – just a single dimension of a single feature – and yet, we found a similar bias away from the prototype. Rather than encode the one feature veridically, participants appear to have encoded a shifted hue. Perhaps the mechanism that underlies the caricature effect is a magnification of the atypical effect we observed through multiple features all moving atypically. While the details of how this one feature case relates to the more complex case of faces is unclear, our results call into question the explanation that we simply encode the distinctive features of a face as they are without the more generic aspects to achieve an exaggerated encoding.

There has been a long debate over whether faces are in some way special in object processing (Farah, 1996; Gauthier & Logothetis, 2000; Kanwisher et al. 1997; McKone et al., 2007; Toveé, 1998). The majority of humans, those without a specific deficit called prosopagnosia, are considered to be experts at facial recognition. In the domain-general expertise explanation of face processing effects, visual object domains other than faces such as cars and birds can be processed in similar ways with experience (Gauthier et al., 2000). Bukach et al. (2006) advocated the use of an expertise framework to understand category specialization. Colors are generally associated with a basic level of categorization (Berlin & Kay, 1969; Rosch, 1975). When forced by a task to make fine subordinate distinctions, a different strategy appears to emerge. Movement in the atypical direction in our experiments was relative to basic categories. Perhaps the fine-grained categorization process is overcompensating for a more natural generalization and homogenization process that occurs when the participant is functioning at the basic category level. The detailed memory of hue demanded of participants in this task was not a typical activity. But for faces, fine-grained distinctions are a basic need. This may indicate that expert processing techniques can be flexibly recruited in real-time to a task and do not depend exclusively on trained distinction making within a domain.

3.6.3 Online Role of Labels

While the prototypical shifts predicted by the representational shift hypothesis were not found and the mechanism underlying the shifts proposed by that hypothesis was not supported, the larger framework of the label-feedback hypothesis (Lupyan, 2012) is not something we are looking to challenge. In the label-feedback hypothesis, language is a pervasive online influence on cognition in the tradition of Whorf (1956). Language is an inherent part of the complex multidimensional system of the normal human adult mind, not something that is switched on and off depending on the task. Instead, labels serve to up-regulate the influence of linguistic knowledge online while verbal interference down-regulates that influence.

One interpretation of the current results would be compatible with the label-feedback hypothesis. Namely, the influence of the labels on memory occurs online and serves to increase the sensitivity of an individual's ability to detect change in the labeled category. In this account, the up-regulation of language's influence in the labeled case allows processing to focus in on the hue resulting in more accurate memory. Sloutsky (2003) discussed the role labels have in directing attention during category learning to relevant similar features among items in a labeled category. Extending the logic from learning itself to the use of learned categories, if labels are features of the category, invoking them will draw attention to the dimension(s) on which the category similarity and distinctions are judged. While labeling did not have the effect of pulling items toward the category prototype in this task, labeling could have a more general modulating influence on encoding.

An alternative explanation could be somewhat consistent with the label-feedback hypothesis but from the opposite direction. The basic categories of color could be essentially automatic (Grill-Spector & Kanwisher, 2005) having language's influence close to ceiling. The preference judgments may serve to distract processing from reaching the level of depth it naturally would regardless of the color label because preference valence needs to be the focus of directed attention (Simons, 2000). Preference judgments would be down-regulating the influence of linguistic category knowledge.

We are agnostic given the present evidence whether labeling has an added effect, the preference judgments have a distracting one, or some combination of the two is at play. Disambiguating the competing interpretations would be an interesting direction for further research. Either way, attention appears to be directed at the relevant dimension for the memory test when colors are labeled while attention is on a different dimension when preferences are being elicited.

3.6.4 Categorical Perception

Our results can also speak to recent developments in the categorical perception literature. Categorical perception is the effect of enhanced discrimination performance when the items being discriminated are across category boundaries. This has been attributed to changes in perception (Harnad, 1987), particularly the enhanced distinctiveness of learned category differences (Goldstone, 1994). Roberson, Hanley, and colleagues (Roberson et al., 2009; Kikutani et al., 2010) proposed a different account suggesting that category labels play a crucial role in categorical perception, with different labels facilitating greater accuracy and faster reaction times.

Hanley and Roberson subsequently updated their account. Conducting a reanalysis of a series of two alternative forced choice categorical perception tasks discriminating between colors or faces (Hanley & Roberson, 2011), they found an asymmetry among the within trials that are traditionally treated as a single condition. On trials where the target item was more typical of the category, or a better exemplar, compared to the foil, participants had a similar proportion correct to between category trials. On trials where the target item was more atypical of the category, a poor exemplar, than the foil, participants performed much worse. These poor exemplar trials account for the overall categorical perception effect. Hanley and Roberson account for this finding through the relative reliability of labels applied to the items. If participants labeled a hue blue when the hue was on its own, the participant who remembers 'blue' rather than the actual color will be more likely to choose the better example of that category - even if the hue they saw was not the best example of the category at test. The items around a boundary are more ambiguous and can be labeled in different ways based on context.

Hendrickson et al. (2012) use a category learning paradigm to investigate the label ambiguity hypothesis (Hanley & Roberson, 2011). They find that there is a pre-categorization asymmetry in addition to the enhanced effect after category learning. If the asymmetry exists prior to learning labels, label ambiguity alone cannot account for the asymmetry. They put forward an account based on unsupervised learning of clusters regardless of labeling.

Our experiments did not contain a classic categorical perception task since we only conducted within category trials. We also did not utilize a two alternative forced choice paradigm. However, the same/different task similarly requires participants to compare their memory for a stimulus to the test items. Rather than use all items as both target and foil, we had set targets with foils in both the typical and atypical direction. Therefore, each target hue was both a good exemplar (atypical trials) and a poor exemplar (typical trials) compared to the current foil. The research above would suggest that there should be enhanced performance on good exemplar, or atypical trials. This is the opposite of what we found. Sensitivity to differences decreased when the test hue was less typical of the category. The label ambiguity hypothesis cannot account for this result.

3.6.5 Conclusion: Memory, Categorization and Reasoning are Intertwined

We examined the effect of active category labeling on hue memory creation. Memory even at 500ms after initial perception is affected by categorical structure, regardless of active labeling. Given the short time scale and the reliable influence of category typicality, it seems safe to conclude that memory and categorization are inextricably intertwined in this task. While our experiments did not look at learning, labels are known to facilitate the learning of categories (Lupyan et al. 2007; Sloutsky &

Fisher, 2012), which is considered to be a reasoning process. In so far as categorization: is based on past experience, is ubiquitous in its influence on memory, and is developed at least in part through reasoning, the historically distinct topics of memory, categorization, and reasoning would appear to be comprised of common elements. As the topics continue to be considered together, the interrelations and underlying processes will become clearer.

Chapter 4

Recognition Memory for Hue: Prototypical Bias and the Role of Labeling[4]

4.1 Introduction

What are the consequences of labeling for perception and memory? Traditionally, language was treated as a separate system within human cognition as it is uniquely human, whereas memory, attention, and perception are held in common with evolutionary ancestors (Chomsky, 1975; Fodor, 1983; Pinker, 1994). However, there is increasing evidence that language uses the same underlying subsystems as other cognitive processes (Perfors, Tenenbaum, & Regier, 2011; Tomasello, 2009). Within such accounts of cognitive processing, interaction between language and memory no longer needs to be the product of separate processes taking place concurrently. Rather, the processes themselves may overlap in their implementation with a high degree of mutual influence.

Linguistic knowledge and non-linguistic knowledge are intimately tied together. To communicate about knowledge, the most pervasive and efficient form of information transfer is through language. Categories are structured knowledge about groups of objects, actions, or whatever phenomena are able to meaningfully hang together. In everyday behavior, category labels are used to refer to phenomena in the world as people communicate with one another and themselves. Most models of categorization (e.g. Heit, 1992; Nosofsky, 2015) do not assign any special influence to the act of applying category labels. The topic has started to garner considerable interest (Blanco & Gureckis, 2013; Hanley & Roberson, 2011; Lupyan, Rakison, & McClelland, 2007; Lupyan, 2008; Richler, Gauthier, & Palmeri, 2011; Richler, Palmeri, & Gauthier, 2013; Roberson & Hanley, 2010). If the act of labeling systematically affects concurrent perception and subsequent memory, it could have pervasive influence on knowledge built out of our experience with the world.

In the present paper we address the issue of how labeling a perceived hue, e.g., calling an orange-red hue "red," affects subsequent memory for that hue, particularly whether there are systematic shifts in memory responses relative to the category typicality structure. We compare labeling to other judgments made at encoding in order to examine the specific effects of labeling on memory for percepts and to examine whether these effects are unique to labeling.

4.1.1 Memory Distortion

Memory is remarkable in its capacity. People are capable of distinctly remembering hundreds of objects (Brady, Konkle, Alvarez, & Oliva, 2008) and scenes (Konkle, Brady, Alvarez, & Oliva, 2010; Shepard, 1967) after relatively brief exposure. At the same time, memory is not necessarily accurate to experience even when a person is very confident of that memory. After hearing an intense verb to describe a car crash - smashed - participants reported remembering broken glass – a marker of a more intense crash – even though there was no broken glass in the video of the crash they had watched (Loftus & Palmer, 1974). Whereas memory is clearly not analogous to video recordings capturing unaltered snippets of audiovisual reality, documented distortions like the eyewitness testimony example have systematic structure. A false alarm – remembering something as old when it is new – in memory for a list of words is more likely if the word is semantically related to the rest of the words in the list (Roediger & McDermott, 1995). The word habitually used to describe the middle color of a traffic light (yellow – “gelb” - in German, and orange – “oranje” - in Dutch) influences the memory of the light's hue (Mitterer, Horschig, Musseler, & Majid, 2009). In situations that are selected for their contrast between reality and description, language is able to influence memory.

How is memory affected by everyday application of language during perception? In situations where language and perception match, does language affect memory for that perception? For instance, as we examine in the present experiments, does labeling a specific hue as "red" change the memory for that hue? A specific hue and the general category of red do not contain the same information. The label "red" applies to the hue, but it also applies to a range of other similar hues. The mismatch between information the label can apply to and the specific item the label is being applied to at a particular point in time could potentially lead to systematic memory distortions. However, that a label can apply to a broad range of category members does not necessarily mean the label in context activates category information with enough strength to alter processing of perceptual input.

4.1.2 How Could Labeling Affect Perception and Memory?

If there were a pervasive influence of labeling on perception and memory, even when the labels agree with the perceptual input, what would this influence be? Here, we explore three alternative hypotheses: (1) labels distort memory towards the label prototype, (2) labels guide the level of specificity needed in the current context, and (3) labels direct attention to a label’s referent among all possible features within a visual scene.

4.1.2.1 Labels Distorting Memory

The representational shift hypothesis (Lupyan, 2008) makes the conjecture that memory is systematically biased towards the category prototype when category labels are active at the same time as when relevant perception is taking place. Specifically, the representational shift hypothesis suggests that perceptual information is perceived equivalently regardless of labeling, then while the information is processed, there is memory distortion from top-down category information being co-activated with the bottom-up perceptual information. The category information is broader than the specific exemplar being perceived. Labeling causes the category information to be more strongly activated than if the category was activated through the perceptual information alone. The general category information and the specific exemplar interact in online processing, shifting the memory of the original percept towards the prototype of the category – the overlapping information between percept and category is more active than idiosyncratic aspects of the current exemplar. It is that shifted representation that is stored in long-term memory. In other words, the orange-red hue labeled as "red" would be remembered as more red. However, this pervasive prototypical distortion of memory based on labeling has been questioned (Blanco & Gureckis, 2013; Richler, et al., 2011, 2013). The representational shift hypothesis, if correct, would represent an accumulating influence of existing knowledge biasing new knowledge towards existing ideas about how the world is structured.

4.1.2.2 Labels Guiding Specificity

Another possible effect is that the label serves as a cue as to what level of encoding specificity is needed. For example, the very question of whether a hue is red or green may suggest that the diagnostic level of detail needed to distinguish the hues for these labels is all that is needed in the current environment, rather than more detailed information such as whether a hue is an orange tinted red or more of a purple tinted red. In this view, an item only needs to be processed sufficiently to be confident the item is in one category and not the other when choosing between two labels, as was the case in related work (Richler, et al., 2011). A superficial orienting task, such as indicating the direction a chair is facing (Blanco & Gureckis, 2013) or the location of the item on the display screen (Richler et al., 2011), could also lead to poor memory for the exemplar. These tasks were designed to influence the strength of a memory through tasks requiring low depth of processing. Under this account, labeling does not have a systematic influence on memory traces so much as simply being a task that may not require specificity.

4.1.2.3 Labels as an Attentional Focus

More broadly, labels may serve as a general cue for attention. When a stimulus or a scene is visually perceived, there are many features initially processed into iconic memory. Guiding the focus of attention to regions within a scene guides the content of accessible working memory for that scene. This was demonstrated in a classic study (Sperling, 1960) by briefly showing participants a display of many rows and columns then immediately cueing a specific row. Participants were able to remember any single row that was cued, but they could not remember the whole array. Focusing attention to a particular row allowed that information to be secured into memory while the perception of the rest of the display was lost. Likewise, using a label concurrently with visual perception would focus the strength of memory encoding towards the features relevant to the label. Under this account, labeling a percept in terms of its color would encourage attention to featural information related to color, but labeling would not be unique in this role, and it should be possible to cue color in other ways as well.

4.1.3 Existing Evidence for Systematic Bias

Labeling bias as discussed here has been investigated in several recent studies using an old/new recognition paradigm (Blanco & Gureckis, 2013; Lupyan, 2008; Richler, et al., 2011). A set of items is presented in a study block, during which time participants are (or are not) asked to label each item. Next, these items and matched lures are presented during a test block. There are challenges associated with using this paradigm to test the representational shift hypothesis, which predicts that memory is shifted systematically towards the prototype (Lupyan, 2008). Poor recognition memory, in the form of misses on an old/new recognition test, need not mean that memory for that item is more prototypical. The memory could be more prototypical, less prototypical, or simply not strong enough to produce recognition regardless of movement in representational space. Returning to our earlier example, if a participant studies an orange-red hue, a failure to later recognize that hue may or may not mean that the representation had shifted towards the red prototype.

In our own prior work (Kelly & Heit, 2014), we had focused on the claim (Lupyan, 2008) that memory shifts during the encoding process. Therefore, we tested immediate memory with narrow study-to-test time scales, 500ms and 5000ms. Memory for hue was tested using the old study items as well as four matched lures that systematically varied relative to the color's prototype. It was the responses on the lure items, either shifted towards or away from the prototype, that were crucial for examining whether and how a representation had shifted. We found that at these tested time scales, memory test responses are systematically biased relative to the category structure, and, surprisingly, are biased away from the prototype.

These shift patterns may not hold at other timescales. It seems possible that, at a timescale of minutes, memory response distortion during recognition would be prototypical rather than atypical. In fact, much of what is known of memory including gist memory (Reyna & Brainerd, 1995) and false recognition of close associates in word lists (Roediger & McDermott, 1995) suggests that memory is broadly influenced by category knowledge and that influence is generally in a prototypical direction.

Additionally, at the brief timescales investigated in Kelly and Heit (2014), there was not a distinctive effect of language. There was a small effect of category labeling that emerged by 5000 ms but was not present at 500 ms. Encoding hue into immediate memory was not apparently affected by online labeling. To determine if online labeling influences processing of hue over time, particularly as memory traces get encoded into long-term memory, here we used a hybrid of our earlier method with a longer study-to-test delay based on the timing used by other researchers (Blanco & Gureckis, 2013; Lupyan, 2008; Richler, et al., 2011).

In related work, Bae, Olkkonen, Allred, and Flombaum (2015) investigated hue estimation (rather than recognition) at 0ms and 900ms, finding varying amounts and direction of bias depending on target hue. However, the Bae et al. experiments did not have a labeling manipulation. Also, Persaud and Hemmer (2014) looked at bias in color recall memory using active labeling immediately before recall, estimating the hue on a color wheel. They found memory to be biased towards the prototype or mean hue value. However, there was no comparison condition without labeling, and labeling itself occurred just before the response. Hence neither of these studies allow us to address questions about the influence of labeling at encoding and whether any effect found is unique to labeling.

4.1.4 Shifts of What Exactly?

The key issues of whether memory truly shifts, and if it does shift, how it shifts relative to the prototype is not discernible from an old/new recognition paradigm with a single matched lure as described above. A different paradigm, one that probes memory at multiple locations, can shed some light on whether the hypothesized shifts are a possibility. If memory test responses are systematically shifted, the role of language in causing those shifts then becomes more interesting. Even showing shifts in memory test responses does not demonstrate that memory itself is shifted – it could be that shifts are introduced at retrieval rather than encoding. However, the mechanism debate is premature without first demonstrating shifts in memory responses.

Yet, mechanisms of shifts in addition to whether shifts exist have already been debated. The representational shift hypothesis (Lupyan, 2008) describes a phenomenon that is specific to overt labeling of percepts. It does not predict that merely knowing a category will cause prototypical shifts of exemplar memory, but that using the label of the category actively changes the processing of perceptual input. This leads to the following prediction: Labeled images will be stored more prototypically and the original image will therefore be less likely to be recognized during a memory test. This would be due to a change in the encoded percept. The specificity of encoding view predicts a different mechanism: less specificity in the memory trace and therefore less recognition of the original labeled images (Blanco & Gureckis, 2013; Richler et al., 2011). The best guess would be a more prototypical guess. Memory probed at multiple locations on a typicality gradient would be predicted to have the same result by both these accounts: labeled items are expected to be more prototypical than un-labeled items. The directed attention account would, in opposition to the other accounts, predict less shifting for labeled items than un-labeled items. This decrease in shifting would be due to more specificity in encoding due to the tested aspect of the item having been highlighted when originally perceived. Shifts in the response for the second and third hypotheses would not necessarily be shifts of the actual percept but could reflect relative susceptibility of a memory trace to the influence of a category prototype at response. For example, a potential mechanism that could underly the directed attention account is described by a prominent model of spatial memory, the Category Adjustment Model, CAM, (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000; Huttenlocher, Hedges, Lourenco, Crawford, & Corrigan, 2007). CAM predicts shifts in spatial memory based on the strength of a memory trace with less specific exemplar memory filled in with category information at retrieval. Applying the principles of CAM to hue memory, memory traces would be encoded as originally perceived but less specific memories should result in more shifted responses while more specific memories would result in less shifted responses.

4.1.5 Experiment Rationale

We conducted a series of four experiments using a hybrid of paradigms from Lupyan (2008), and Kelly and Heit (2014). The aim of the work was to investigate whether there is a systematic bias predicated on labeling categories when encountering a percept. To give a preview, participants were presented with colored silhouettes and made two types of judgments about them (with different pairs of judgments in each experiment; see Table 4.1). Afterwards they were given a five-alternative forced choice memory test for the hue of the silhouette among slightly different hues; see Figure 4.1 for an example display. Experiments 1 and 2 established that shifts were taking place. Additionally, the experiments investigated whether overt labeling of categories leads to a different pattern of shifts than a comparison condition, in Experiment 1 with the preference judgment comparison condition as in Lupyan (2008), then in Experiment 2 with another categorization task, an animacy comparison condition. Experiments 3 and 4 explored the role of language in shifts. Experiment 3 explored whether the effect of language is unique and Experiment 4 explored whether overt color labeling is required to produce the effect. This sequence of experiments allowed us to test for systematic shifts relative to category structure, to determine the direction of those shifts, and to examine the role of online labeling in producing shifts at the timescale of about five minutes between study and test.

4.2 Experiment 1

To look for shifts in memory responses after a few minutes of processing, we employed the basic design of the original representational shift paradigm (Lupyan, 2008): a study phase during which participants viewed target items and made categorization or preference judgments about them, immediately followed by a surprise test phase where participants identified which items they had seen previously. Rather than display targets and a single matched foil in isolation at test, we tested participants on an array of five hues: the original hue along with two hues more typical and two hues less typical of the color category. The richer false alarm possibilities make the direction and magnitude of a response shift observable.

In line with Kelly and Heit (2014), we used color categories rather than object categories, such as chairs and lamps, used in the studies by other researchers because color has a better-defined representational space. We focused on the single dimension of hue, leaving two directions of possible shift: away from or towards the prototype of a color category. Color is a domain where memory has, in some cases, proven accurate to experience (Perez-Carpinell, Baldovi, de Fez, & Castro, 1998), and is a domain where categories affect perception (Kay & Kempton, 1984; Winawer et al., 2007). The possible veracity of perception coupled with the demonstrated distortion of perception in different settings make color an ideal testing ground for whether overt language can cause systematic memory distortions.

Experiment 1 began to address both of our aims: establishing whether there are directional shifts in memory test responses and, should shifts be found, investigating the role of language in those shifts.

4.2.1 Method

4.2.1.1 Participants

Thirty-eight undergraduate students were recruited from UC Merced's participant pool.[5] All participants were monolingual or early (by age 10) bilingual English speakers. They had not otherwise participated in color-based experiments in our lab. Participants had normal or corrected to normal vision. They were also tested to have normal color vision, using the CITY test (City University, 2002) at the conclusion of the experimental session.

4.2.1.2 Stimuli

Forty silhouettes were created in Adobe Photoshop for the study phase. Twenty silhouettes were living things such as a giraffe and a butterfly, and 20 silhouettes were non-living things such as a pan and an airplane. Eight main target hues, 4 reds and 4 greens, were selected. The selected colors had the lightness and saturation values of their category's focal color (Sturges & Whitfield, 1995). The hues were evenly distanced from each other within hue space centered on the category focal color hue. The semi-random (see Appendix A for details) creation of colored silhouettes preserved color and animacy balances while counterbalancing the color/shape pairings across participants. The animacy features of the

---

[5] A preliminary version of this experiment was presented in a conference proceedings paper, Kelly and Heit (2013). We have omitted one participant due to missing information on native-English speaking status.

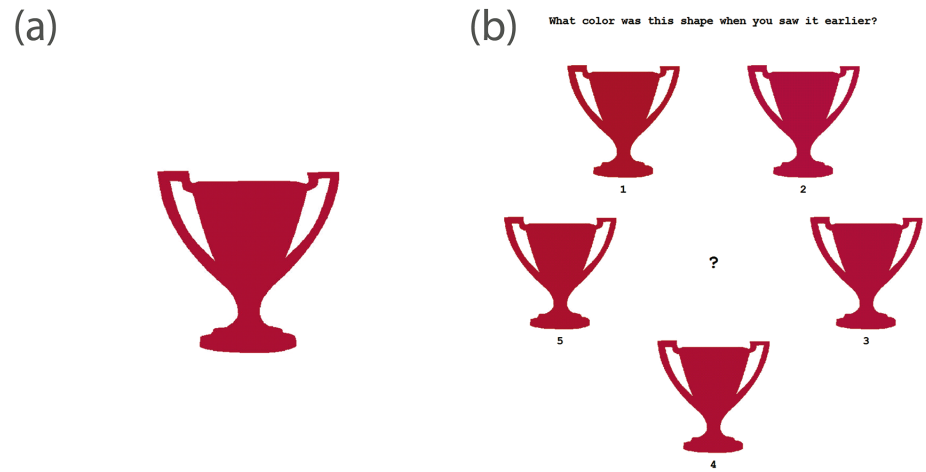

Figure 4.1. Stimulus presentation. (a) During the study phase, participants saw a single silhouette as they made a judgment. (b) During the test phase, participants selected from an array of 5 hues for a particular silhouette shape. The question reads, "What color was this shape when you saw it earlier?"

silhouettes were not required in the current experiment but were necessary for an animacy judgment condition used in Experiments 2-4. Six additional colored silhouettes were developed for practice trials, using the colors yellow and purple as well as different shapes than those used in the main task.

For the test phase, four variations of each of the 8 target hues were created at intervals of 2° along the hue dimension in CIE L*ab color space (see Appendix A for details). This resulted in an evenly spaced test scale for each silhouette of two hues closer to the prototype, the original hue, and two hues further from the prototype. The typicality of the hues was empirically tested in a norming experiment using an independent participant sample, which is reported in Appendix B.

4.2.1.3 Procedure

The experiment consisted of two main parts: a study phase and a test phase. Each participant encountered 6 practice study trials, then 80 study trials followed by 40 test trials.

For the study phase, participants were instructed in the two types of judgments: color categorization, "What color is this?" ("1" and "2" for red and green, with color/key assignments counterbalanced across participants) and preference, "Do you like this?" ("3" for like or "4" for dislike), the conditions used in Lupyan (2008). The trials were presented in 8 alternating judgment blocks (e.g., a block of red/green, followed by a block of like/dislike), 10 trials per block, with the starting condition counterbalanced. Each silhouette was judged twice by each participant for the same judgment type (e.g., if the giraffe was judged for color the first time it appeared, the giraffe would again be judged for color the second time it appeared). Each silhouette was judged in each condition, between

Table 4.1

Proportions of Responses by Typicality Level at Test

	Condition	Least Typical	Less Typical	Original	More Typical	Most Typical
Exp. 1	Color	0.155	0.155	0.191	0.236	0.263
	Preference	0.127	0.159	0.173	0.234	0.307
Exp. 2	Color	0.170	0.163	0.171	0.223	0.273
	Animacy	0.125	0.143	0.177	0.202	0.354
Exp. 3	Color	0.182	0.132	0.177	0.208	0.302
	Animacy	0.192	0.143	0.150	0.240	0.275
Exp. 4	Preference	0.170	0.127	0.181	0.192	0.330
	Animacy	0.165	0.146	0.180	0.208	0.301

participants. The participants had been instructed to remember the silhouettes as they would show up more than once but were not explicitly told of a memory test. Each trial consisted of a fixation cross (1500 ms), the silhouette to be judged (300 ms), a question mark eliciting the judgment for that silhouette (up to 700 ms), and a blank screen (up to 1000 ms). The trial ended after a response was given. If the trial timed out a "Too Slow" message briefly appeared on the screen before a new trial started. The timing and block design of the study phase closely followed the procedure of Lupyan (2008) Experiment 1.

After the study phase, participants were then tested for hue memory. The memory trials consisted of a circular array of the 5 hue variants of a particular silhouette (Figure 4.1). The array consisted of the hues in graded clockwise order with the most typical hue rotated to a random position on each trial, resulting in a consistent appearance of selecting from a gradient of hues but avoiding position effects that would be present in a line. Each of the 5 positions had a location label 1 through 5 that participants entered on the keyboard to make a selection. There was no time limit imposed on the memory test responses, with an inter-trial interval of 1500ms.

4.2.2 Results and Discussion

In spite of screening for color vision deficiencies, 4 participants were excluded for low (<80%) color categorization accuracy. Six additional participants were excluded for response patterns suggesting a lack of engagement such as always pressing the same response key. As a result, 28 participants were included in the analyses. The same pattern of results was maintained when all 38 participants were analyzed. The included participants had a color categorization accuracy of 90.56%.[6]

The first consideration was whether there were the hypothesized systematic shifts. Responses on the memory test were transformed to a consistent -2 to +2 scale, with -2 corresponding to choosing the item 4° of

[6] No individual hues were more accurately categorized than the others (See Appendix A for details).

hue space more distant from the prototype than the original, 0 choosing the original item, and +2 choosing the item 4° closer to the prototype than the original. The proportion of responses for each of these hue types by condition are reported in Table 4.1.

Participants showed a systematically biased memory for hue toward the prototype. We conducted two-tailed, one-sample t-tests comparing the mean typicality of responses for each judgment type against the value 0 representing the original hues. The color categorization mean was 0.29 (standard deviation 0.42) which is significantly more prototypical than the original hues, $t(27)=3.75$, $p<0.001$, Cohen's $d=0.71$, $BF_{10}=38.97$. Here and in subsequent reports, the Bayes factor, calculated using JASP (Love et al., 2015), represents the evidence of the alternative hypothesis (H_1) against the null hypothesis (H_0), with values above 1 being evidence for H_1 and values between 1 and 0 being evidence for H_0. To interpret the current value, a BF_{10} of 30-100 is considered very strong support for the alternative hypothesis (Jeffreys, 1961). The preference judgment condition mean was 0.44 (standard deviation of 0.42) which is also significantly more prototypical that the original hues, $t(27)=5.57$, $p<0.001$, $d=1.05$, $BF_{10}>1000$. A Bayes factor above 100 is considered decisive support for the alternative hypothesis. Here, we provide direct evidence for shifts toward the prototype at a study to test delay of approximately 5 minutes. In an effort to verify the assumptions of these analyses, we examined the response distributions for both color judgments and preference judgments. Each distribution failed a test for non-normality, the Shapiro-Wilk test, p-values >0.05. The distributions were each not statistically different from a normal distribution.

This design did not have a measure of hue/shape pair memory. The memory test does not distinguish random guessing due to no memory for the hue when a shape is displayed at test and true biased memory. We conducted a control experiment reported in Appendix B. Participants in that experiment had a 67% accuracy rate of categorizing a gray silhouette cue as having been red or green at study – effectively a cued recall task. Participants clearly have some memory of the pairings but it is not at ceiling. To determine whether there were differences in accurate hue memory across the encoding tasks, we analyzed the proportion of objectively correct responses (the Original column in Table 4.1) by condition. There was no difference in these proportions (paired, two-tailed $t(27)=0.92$, $p>0.05$, $d=0.17$, $BF_{10}=0.29$), indicating that the conditions contained equal original hue responses with any differences between conditions being within the pattern of the lure responses. Across all four experiments, the findings illustrate that simply looking at these accurate responses to the original hue is not as informative as looking at the whole pattern of false alarms.

In order to investigate potential condition differences, we tested the mean response typicality of the labeling and preference conditions against each other using a two-tailed, paired-sample t-test. The difference between conditions (label condition mean: 0.29 and preference condition mean: 0.44) was not significant, $t(27)=1.97$, $p=0.059$, $d=0.34$, $BF_{10}=1.08$. With a

Bayes factor just above 1, a difference between the conditions is only weakly evident. The null hypothesis of no difference is not supported, leaving the condition difference inconclusive. A chi-square analysis comparing the distribution of responses at each typicality level between the two conditions was also non-significant, $\chi^2(4, N=1160)=3.51$, $p>0.05$. Since the t-test yielded a p-value close to the significance level of 0.05 and with an inconclusive Bayes factor, it is natural to speculate as to whether there is a difference that we do not have enough power to detect. However, note that the shift was actually weaker for the labeling condition, the opposite pattern of results from those proposed by the representational shift hypothesis (Lupyan, 2008). If this pattern holds, color labeling would actually lead to weaker rather than stronger shifts in comparison to another kind of judgment. Moreover, labeling does not uniquely cause shifted response patterns. This finding cannot support the specific formulation of memory distortions hypothesized by the representational shift hypothesis. The other possibilities of labels guiding specificity or focusing attention could be compatible with the demonstrated systematic bias.

Finally, we collected reaction times (RTs) for both judgment responses and memory test responses with measurement beginning at stimulus presentation - either the single item in the study phase or the array of items in the test phase. The color judgment RTs (median 425.62ms) were significantly faster than the preference judgment RTs (median 520.10ms) according to the Wilcoxon Signed-Rank Test ($V=71$, $p<0.002$) which is appropriate for comparing paired sample non-normal distributions. This replicates a reaction time difference reported in Lupyan (2008). The additional processing time needed for preference judgments has been taken as indicative of more depth of processing in the preference condition (Blanco & Gureckis, 2013; Lupyan (2008); Richler et al., 2011, 2013). There was no significant difference for color memory RTs (median 3444.20ms) and preference memory RTs (median 3982.33ms), $V=185$, $p=0.695$.

Overall, the main finding of Experiment 1 was that responses shift relative to category structure at the retention interval of a few minutes. This happened regardless of the orienting task being color labeling or preference judgment. Additionally, there was not a significant difference between the two orienting conditions. The pattern of results, though not conclusive, lean toward less bias for labeled hues, the opposite pattern from what the representational shift hypothesis would predict. The possible roles of labels setting a minimum level of specificity and of labels focusing attention to particular features are not distinguished by Experiment 1. Preference judgments are not categorical and specific in the same way as category label judgments are. The level of specificity needed to make a preference judgment is less clear than discriminating between two well-defined categories in the color label condition. Which features preference judgments cause participants to focus on is unclear and likely subject to individual differences with some participants focusing on liking or disliking the color, the shape, the animal or object concept indicated by the shape, or a holistic view of more than one of these dimensions. An alternative contrast condition which is a more defined categorical task would help to clarify these issues.

4.3 Experiment 2

In Experiment 1, we demonstrated that there are prototypical shifts using a serial study/test paradigm. To replicate the prototypical shifts and explore potential orienting task differences further, Experiment 2 was similar to Experiment 1 but replaced the preference judgment condition with an animacy judgment task. Judging the animacy (living or non-living) status of a silhouette is a categorization task like the color judgment. Responding to animacy is a task which people – even infants – can perform easily (Mandler & Bauer, 1988).

The key information from the stimuli for the animacy task is the shape of the silhouette rather than the color. In preference judgments, it is possible that color plays a significant role in the judgments. Liking or disliking an item is a holistic judgment which would likely take into account the combination of the hue and shape. The separation of key information – hue for color judgments, shape for animacy judgments – for the two orienting tasks helps increase the difference between tasks. The minimum specificity alternative would suggest little to no difference between conditions: distinguishing between red and green takes little specificity of hue processing and distinguishing between living and non-living shapes requires no specificity of hue processing. The directed attention alternative would predict higher accuracy for hue in the color judgment task, where hue is highlighted over shape. The results of this experiment can confirm the pattern found in Experiment 1 which is inconsistent with the representational shift hypothesis and lay groundwork for distinguishing between specificity and directed attention. Although this experiment alone cannot conclusively distinguish between the specificity and directed attention accounts, a difference between conditions would be more easily explained by the directing attention account.

4.3.1 Method

Forty undergraduate students were recruited as in the previous experiment. Participants answered the judgment questions "What color is this?" and "Is this a living thing?" in alternating blocks during the study phase. Otherwise, the stimuli and procedure were the same as in Experiment 1.

4.3.2 Results and Discussion

Twelve participants were excluded for low (<80%) color accuracy, low (<70%) animacy accuracy, or response patterns such as always pressing the same response key. Hence, 28 participants were included in the analyses reported, although the conclusions are the same based on data from all the participants. The included participants had 94.38% color accuracy and 88.84% animacy accuracy.

Participants showed systematic biased memory for hue in both the label and animacy conditions. The prototypical shifts observed in Experiment 1 were observed again. The color condition mean was 0.27 (standard deviation of 0.50) which is significantly more prototypical than the original hue value 0, $t(27)= 2.86$, $p<0.01$, Cohen's $d=0.54$, $BF_{10}=5.54$. The animacy condition mean was 0.52 (standard deviation of 0.36) which is also significantly more prototypical than 0, $t(27)=7.54$, $p<0.001$, $d=1.43$, $BF_{10}>1000$. Both distributions failed the Shapiro-Wilk test, with both p-values >0.05. Comparing the proportion of accurate responses (Original column, Table 4.1), there was no difference by condition (paired, two-tailed $t(27)=0.20$, $p>.05$, $d=0.04$, $BF_{10}=0.20$). Again, there was no evidence that any differences by condition are due to having more memory for hue/shape pairings in one condition over the other.

We again found color labeling does not uniquely cause response shifts, and further, it leads to weaker shifts than another kind of judgment. The animacy mean response typicality 0.52 was statistically different from the labeling mean response typicality 0.27, $t(27)=3.12$, $p<0.01$, $d=0.59$, $BF_{10}=9.53$, with animacy-judged hues shifting more strongly in a prototypical direction than color-judged hues. The Bayes factor between 3-10 indicates substantial support for a difference (Jeffreys, 1961). A chi-square analysis comparing the distribution of responses at each typicality level between the two conditions (see Table 4.1 for response rates) was also significant, $\chi^2(4, N=1120)=10.916$, $p<0.05$. These results point to different patterns of bias for the two encoding conditions with labeled hues being less biased towards the prototype than non-labeled hues.

Reaction times (RTs) followed the same pattern as in Experiment 1. Color judgment RTs were significantly faster (median: 444.71) than animacy judgment RTs (median: 528.93) according to a Wilcoxon Signed Rank Test, $V=0$, $p<0.001$. Whereas animacy is a categorization task like color, the categorizations are at different hierarchical levels with color at a basic level and animacy at a superordinate level. The longer RTs for superordinate categorizations compared to basic categorizations is expected (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). There was no statistical difference between the conditions for memory test RTs, $V=142$, $p=0.171$ (color median: 4221.55, animacy median: 4677.50).

Experiment 2 replicated the prototypical shift finding from Experiment 1 using an alternative orienting task. Responses on a memory test for hue systematically shift at this timescale. Again, both orienting tasks tested led to bias towards the prototype. Here the strength of shifts between conditions was significantly different with animacy judgments leading to more biased memory than color labeling. The prediction of the representational shift hypothesis, that overt labeling strengthens category prototype information during encoding, was again not supported. Still, this experiment does not conclusively distinguish between the directing attention and specificity hypotheses for the role of labeling. The condition difference here could be due to attention focusing on the color and separately the shape in the color and animacy judgment tasks respectively. Alternatively, labeling color could set a minimum specificity of encoding that was greater than for the animacy condition for which the hue was irrelevant. In Experiment 3, we employed a non-label manipulation to further distinguish between the alternative accounts.

4.4 Experiment 3

Why would labeled hues have a tendency for weaker shifts than in conditions without labeling? One possible explanation is that the labeled feature is processed to a minimum specificity for distinguishing between the alternatives. Since the labels used here (green or red; living or non-living) are broad category labels, this could result in lower accuracy than if a more stringent sub-category criteria were applied. Another possibility is that labels direct attention to the appropriate features causing hue to be remembered with greater accuracy in the color labeling condition.

In Experiment 3, participants were informed of the upcoming memory test, including being shown a display of how narrow the differences between hue options are at test. By informing participants of the hue memory test prior to the orienting tasks and showing them how highly specific the hue test would be, the knowledge of this test could define the focus of attention and the specificity of encoding during the study phase of the experiment. If the effect of labeling is actually an effect of attention to highlighted features, the higher accuracy found in the color condition in Experiment 2 should be achievable in the animacy condition if attention were directed to the hue feature by an alternate means. Therefore, both orienting conditions should produce weak shifts. If instead labeling sets the specificity of processing in the relevant dimension, alerting participants of a need for greater specificity than distinguishing between the label options would result in less shifting than the color labeling condition in Experiment 2.

Additionally, this manipulation of the instructions allows the question of labeling uniquely decreasing shifts to be examined. Eliminating the difference between the labeling and animacy conditions with an instructional (that is, a non-label-based) manipulation would suggest that labeling is not the only way to increase hue memory accuracy.

4.4.1 Method

Thirty-eight undergraduate students were recruited as in the previous experiments. The initial instructions, both presented verbally from the experimenter and written on-screen, informed participants of the hue memory test that would follow the judgment task. The on-screen instructions displayed an example test question in order for participants to experience the small variation in hues displayed at test. Otherwise, the stimuli and procedure were the same as in Experiment 2 using the categorization and animacy conditions.

4.4.2 Results and Discussion

Eight participants were excluded using the same criteria as Experiments 1 and 2. Therefore, 30 participants were included in the analyses. Excluding participants did not change the pattern of results. The remaining participants had a color categorization accuracy of 91.42% and an animacy categorization accuracy of 88.00%.

We again found evidence of systematic prototypical shifts in the memory test response patterns for both the labeling and animacy conditions. The condition means were again significantly shifted toward the prototype (one sample, two-tailed t-tests against original hue value of 0: color mean = 0.32 (SD = 0.43), $t(29)=4.07$, $p<0.001$, $d=0.74$, $BF_{10}=88.57$; animacy mean = 0.26 (SD = 0.43), $t(29)=3.36$, $p<0.01$, $d=0.61$, $BF_{10}=16.41$). Shapiro-Wilk tests did not indicate non-normal distributions, p-values > 0.05. Comparing the original hue response rate by condition again yielded a null result (paired, two-tailed $t(29)=1.21$, $p>0.05$, $d=0.22$, $BF_{10}=0.38$).

Contrary to the possibility that the mitigating effect of labeling is unique to language, the conditions were not significantly different from each other (two-tailed, paired-sample t-test: $t(29)=0.63$, $p>.05$, $d=0.12$, $BF_{10}=0.23$; chi-square test: $\chi^2(4, N=1200)=3.845$, $p>0.05$; see Table 4.1 for response rates). The inverse Bayes factor (BF_{01}) is 4.27, which is considered to be substantial evidence for the null hypothesis that the conditions are not different. Forewarning the participants of the memory test affected participants' processing, seemingly in a similar way to the effect of labeling.

In Figure 4.2, it is clear that the mean response typicality for both conditions in this experiment mirror the color categorization condition mean response typicality in Experiment 2. This pattern was verified as statistically significant by conducting a 2 x 2 ANOVA with judgment condition within-subjects and instructional condition between-subjects (across experiments). There was an interaction of judgment and instructional conditions, $F(1,56)=6.752$, $p<.02$, $\eta^2=.10$, $BF_{Forward}=4.08$. The Bayesian analysis for the full model of both main effects and the interaction was inconclusive at $BF_{10}=1.01$, but the interaction term $BF_{Forward}$ of 4.08 indicates that just the interaction, not the full model, has substantial support against the null. This suggests that the instructional manipulation selectively and significantly changed the average typicality of the selected hue in the animacy judgment condition while not having much of an effect on the color labeling condition. Although we acknowledge the possible limitations of cross-experiment statistical comparisons, we note that participants for the two experiments were drawn from the same pool and recruited in the same manner. This pattern is more consistent with the hypothesis that labeling directs attention to the relevant feature rather than setting a minimum specificity of processing.

Reaction times again followed the same pattern as Experiment 1 and 2 with significantly different RTs for the judgment responses (color median: 470.68, animacy median: 549.69, Wilcoxon Signed Rank Test, V=6,

$p<0.001$) and no statistical difference in RTs for the memory test responses (color median: 3840.95ms, animacy median: 3920.65, Wilcoxon Signed Rank Test, $V=247$, $p=0.777$). Reaction time differences have been taken as a proxy for depth of processing, and the repeated longer processing time for animacy would indicate more depth of processing. However, the combination of reaction time differences and a lack of difference in memory bias across conditions suggest there is not a linear relationship between length of processing and bias at least as is relevant to hue memory. Either processing time is not a good measure of depth of processing or depth of processing is not explanatory for the memory bias results.

Telling participants to attend to hue decreases hue memory response bias in this paradigm. The difference between orienting tasks disappears with both conditions leading to more accurate memory for hue. This lends support to the idea that rather than overt labeling selectively recruiting category knowledge in this paradigm, labeling acts as a cue to encode the hue feature of the exemplar with more precision. When making an animacy judgment the shape of the stimulus was most important for the task. Participants demonstrated they could, under suitable task instructions, attend to both hue and shape by maintaining a similar level of animacy judgment accuracy while also increasing their accuracy in hue memory.

4.5 Experiment 4

Finally, we put the preference and animacy judgment conditions in the same experiment leaving out color labeling altogether. It has been suggested that language is a component of conceptual processing that can be up-regulated or down-regulated based on task demands (Lupyan, 2012). Specifically, engaging the language network for a task can globally recruit language related processing to a task. Labeling color specifically enhanced hue memory response accuracy, decreasing shifting relative to non-labeled stimuli in Experiments 1 and 2. Color labeling could have a specific effect of increased accuracy on labeling trials but could also have a global effect of priming color category knowledge driving the overall shifts observed in all experiments and conditions. If color labels are not overtly engaged, will systematic response shifts still occur? If categorization affects perception only when relevant linguistic categories are explicitly engaged, it would suggest a loose coupling of perception with conceptual effects led by language. If categorization affects perception regardless of linguistic engagement, it would suggest a stronger coupling of perception with conceptual information. Preference judgments and animacy judgments do not require participants to attend to the colors and their labels. The color and silhouette pairs were chosen to be arbitrary and not semantically indicative of each other. These conditions without a color labeling condition and without foreknowledge of a color memory test provide a context that does not require participants to attend to hue during the study portion of the experiment.

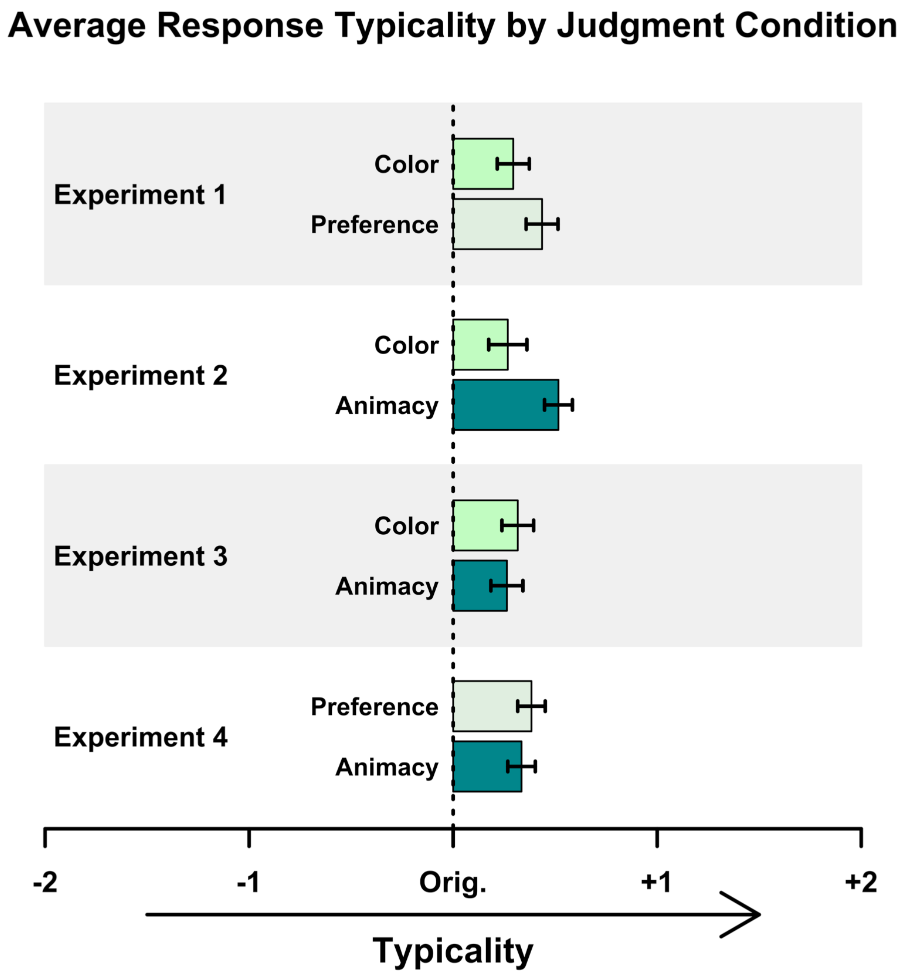

Figure 4.2. Mean typicality of response by condition. The dotted line represents the hues originally shown to the participant while the bars indicate the average typicality of the items with more typical hues to the right. This orientation visually displays the magnitude of the response shifts in hue typicality space.

4.5.1 Method

Forty undergraduate students were recruited as in the previous experiments. Participants answered the judgment questions "Is this a living thing?" and "Do you like this?" omitting the color categorization condition. Otherwise, the stimuli and procedure were the same as in Experiment 1 and 2 with the hue memory test again a surprise following study.

4.5.2 Results and Discussion

Three participants had low animacy accuracy (<70%) and were excluded from the analysis leaving a sample size of 37. Excluding these participants did not change the pattern of results. Included participants had an animacy accuracy of 89.93%.

We again found evidence for systematic prototypical response shifts. The response means for each condition were both significantly different from the target response (two-tailed, one-sample t-tests against the original hue value of 0: preference mean = 0.38 (SD = 0.42), $t(36)=5.69$, $p<0.001$, $d=0.94$, $BF_{10}>1000$; animacy mean = 0.34 (SD = 0.43), $t(36)=4.97$, $p<0.001$, $d=0.82$), $BF_{10}>1000$. The response distributions for both conditions again were not significantly different from normal according to the Shapiro-Wilk normality test $p>0.05$. There was no difference by condition for the objectively correct original hue response rate at test (paired, two-tailed $t(36)=0.06$, $p>0.05$, $d=0.01$, $BF_{10}=0.18$).

The mean typicality of response for the two conditions were not different from each other (paired sample, two-tailed t-test: $t(36)=0.68$, $p>0.05$, $d=0.11$, $BF_{10}=0.22$; chi-square test: $\chi^2(4, N=1120)=1.294$, $p>0.05$, see Table 4.1 for accuracy rates). Again, looking at the inverse Bayes factor, $BF_{01}=4.55$, there is substantial evidence for the null hypothesis that there is no difference in magnitude of prototypical shifts between the two encoding conditions.

The reaction times tell a similar story to the previous experiments. There was a significant difference between the more quickly responded to animacy condition (median: 523.36ms) and the preference condition (median: 560.82ms) according to the Wilcoxon Signed Rank Test, V=518, $p<0.02$. This provides evidence that there is more depth of processing in preference judgments than animacy judgments. There was no significant difference in memory test RTs (animacy median: 4099.75ms, preference median: 3813.05ms, Wilcoxon Signed Rank Test, V=338, $p=0.846$).

With no directing of attention to color either through instructions or through an overt labeling condition, participants did not have any indication that paying attention to color was important for a future task. The results of the memory test in this case continued to show systematic shifts of hue memory responses towards the prototype. Overt language was not required to obtain shifts relative to the category structure. The lack of an effect of condition is consistent with these encoding conditions equally not calling attention to hue. The shifts occurring regardless of color labeling indicates a strong influence of category knowledge in this task with overt priming of the labels not needed.

4.6 General Discussion

In these experiments we set out to investigate (1) if recognition memory for hue is systematically shifted toward the prototype, (2) how overt labeling of colors affects the directional shifts and (3) if any effect of overt labeling on the shifts is unique. These questions were specifically investigated within a context of labels matching their referents to expand beyond known influences of language on memory when the language is contrastive.

Consistently, we have demonstrated that there is systematic bias to responses on a five-alternative forced-choice recognition memory test. Participants on average chose hues that were more prototypical of the basic color category than the hue that they had originally seen in all conditions in all four experiments. This result replicates the existence of response shifts we had previously demonstrated (Kelly & Heit, 2014). However, at the present timescale tested of a few minutes, the direction of the shifts was towards the prototype rather than away from the prototype as had been seen at the previous 500 and 5000 millisecond timescales.

The present findings are novel evidence of prototypical shifts in hue memory test responses. By demonstrating the existence of prototypical response shifts at this time scale in all conditions, the results provide strong evidence of a systematic influence of categorical knowledge on hue recognition memory. Whether the prototypical response shifts are a demonstration of true shifts of encoded memory traces, or if they are due to response bias after memory is retrieved is still up for debate, a debate our additional results inform. Our further analyses and manipulations from experiment to experiment began to investigate the role of labeling in influencing the response shifts.

With regard to the role of language, we found that labeling does not produce stronger typical shifts than comparison conditions. Participants chose more typical hues at test when they had made preference or animacy judgments about a silhouette than when they had labeled the color category. The results do not support a special role of overt labels as top-down category prototypical attractors for memory traces prior to encoding as proposed by the representational shift hypothesis (Lupyan, 2008). Instead, labeling color categories had the opposite effect decreasing the magnitude of the shifts relative to other orienting tasks. This pattern was also not predicted by the specificity hypothesis, which would predict weak memory traces susceptible to the category information at response.

We followed up these findings by testing whether the label effect was driven by the third hypothesis, labels directing attention, while also more directly testing the specificity hypothesis. Our manipulations, informing participants of the upcoming memory test and showing participants an example test display, separated the minimum specificity needed by the task and the directing of attention to hue from the act of labeling (Experiment 3). Participants knew they needed to attend closely to hue with a high level of specificity and map that knowledge onto the shape cues. The manipulations eliminated the difference between conditions, and the response shifts in the two foreknowledge conditions of Experiment 3 were consistent with the color labeling condition of Experiment 2. Foreknowledge improved accuracy in the animacy condition but did not alter performance in the color labeling condition. This is consistent with the directing attention explanation of the label effect – participants when labeling color pay more attention than they would when the shape is diagnostic for their labeling task resulting in more accurate memory. If the specificity of processing idea were true, we would expect to have seen an improvement in hue recognition accuracy for the color labeling condition as well since the example display makes clear much more specific knowledge of the hue is needed in the memory task than simply differentiation between green and red.

Additionally, Experiment 3 speaks against the idea that labeling has a unique effect on hue memory response shifts. While the previous experiments clearly did not support the idea of labels as prototypical attractors magnifying directional shifts, they could indicate labels uniquely decrease shifts. When given foreknowledge of the hue memory test, participants showed reduced shifting in the animacy condition matching the memory performance of the color labeling condition. This evidence speaks against the general hypothesis that labeling causes unique changes in the memory responses.

Finally, we tested whether explicit labeling during the task was required for the response shifts (Experiment 4). Lupyan (2012) suggested the influence of language can be up-regulated or down-regulated in the overall perceptual processing network. In terms of the present experimental design, color labeling in one condition would attune the system to the labeled categories and produce the overall category-relative shifting effect regardless of the trial condition. As mentioned earlier, hue memory has been shown to be both accurate (Perez-Carpinell et al., 1998) and potentially distorted (Kay & Kempton, 1984; Winawer et al., 2007). We conducted a final experiment using the same paradigm comparing the two non-labeling conditions with no overt mentions of the hue color labels and with the hue memory test a surprise. There continued to be an overall prototypical bias in hue memory indicating that explicit labeling during the task is not required to produce the effect.

Together, the results show a pervasive, systematic influence of category knowledge during a hue recognition memory test resulting in prototypical response shifts. Further, category-consistent labeling has an effect of decreasing bias, is not unique in producing the decreased effect, and is not required to produce category-relative response shifts. Labeling appears to efficiently direct attention to a particular feature of a stimulus relevant to that label resulting in increased memory accuracy for that feature.

4.6.1 Shifts of What?

As can be seen in Table 4.1, in no condition across all four experiments did participants choose the original hue at even chance levels. Four lures and the original hue at test means the percentage each item would be chosen at random was 20%. In all conditions, the most typical lure, followed by the second most typical lure was chosen most often. They were the only items chosen above chance.

As mentioned in the introduction, there are two possible interpretations of the response shifts. One possibility is memory traces themselves shift. While they do not shift only, or more strongly, when labeled, it is possible that all of the representations are shifting towards the prototype. In this interpretation, labeling the color category reduces the amount of shift of the memory trace by calling attention to the color feature of the silhouettes during initial encoding. Encoded memory would be influenced by the prototypical category information, with the labeled hues less influenced. The other possibility is memory is not shifting so much as there is a strong typical response bias that can overwhelm the memory trace. With no memory of the hue, or a memory not specific enough to distinguish between the hue options at test, participants tend to choose the hue that is most typical of the color category, which is also the central tendency of all hues of that color category presented during the experiment. Within this interpretation, labeled hues would have a stronger memory trace and would be less subject to the influence of the category during retrieval. In both cases, memory is improved rather than distorted by the act of labeling during encoding.

4.6.2 Limitations

Although our results of bias and mitigation of bias are notable within the scope of our investigation – hue space with basic color labels – we acknowledge that this does not necessarily translate to pervasive bias in perception and memory during everyday life. The current experiments constituted a laboratory-based investigation with some unnatural controls. Colors usually vary in brightness and saturation in addition to hue. They are normally seen under different lighting conditions. Objects will vary in dimensions other than shape and hue. Therefore, natural contexts and situations might not result in the same shifts we observed above.

For our paradigm, color space was a superior domain in comparison to object space due to its well-defined parameters and known average focal or prototypical coordinates for the basic color categories in English. However, color is one feature rather than a whole object and may behave differently. Additionally, since color belongs to the class of categories that have cross-boundary categorical perception effects, shifts may happen more readily or may be more exaggerated than in other potential test spaces.

Labeling itself is a somewhat unnatural task. Not much time is spent overtly labeling the objects and features encountered in everyday life. More often labels are used as part of a broader discourse about the context speakers find themselves within. So while labeling has an effect of minimizing bias with the current paradigm one could imagine with a different paradigm language having less of an attention directing effect. How language affects perceptual processing in natural perception of the world is not directly addressed by our controlled laboratory-based task.

Acknowledging these limitations, our aim with this set of experiments was to address the specific question of memory accuracy when labels are applied continuing a recent line of inquiry (Blanco & Gureckis, 2013; Kelly & Heit, 2014; Lupyan, 2008; Richler et al. 2011). As discussed below and in the existing literature, the demonstrated shifts have theoretical implications that reach beyond this isolated laboratory task.

4.6.3 Prototypical vs. Atypical Shifts

Kelly and Heit (2014) showed atypical shifts for all items in all conditions using the same hues as were used in the present experiments. The experiments focused on testing the prediction of prototypical shifts occurring during initial processing (Lupyan, 2008). In Kelly & Heit (2014), the memory test was an immediate test after the encoding task at 500ms or 5000ms with the encoding and test tasks interwoven. Participants were asked if the just previously displayed and judged hue, using color labeling and preference judgment conditions, was the same or different from the subsequent hue displayed while the participant responded. Kelly and Heit (2014) showed that shifts are a measurable phenomenon and that there was not evidence for the predicted prototypical shifts at this short timescale. It did not clarify whether shifts were exclusively atypical or what the role for labeling would be in a long-term memory task.

In the present experiments, we used a hybrid of the original paradigm from Lupyan (2008) and the paradigm we developed to detect shifts. We continued to use the domain of hue and kept the four lures rather than one, but reverted to the serial study block followed by test block structure. The present experiments were meant to test for prototypical shifts at a timescale similar to those used in studies finding a label effect and to investigate the role of language. Using this hybrid paradigm, prototypical shifts were evident.

We have not yet experimentally investigated the relationship between these two sets of results. It could be that one of the above parameters or task demands accounts for the difference in shift direction. We favor the longer time delay and the addition of an intervening task in the present experiments as the factors driving the differences. Kelly and Heit (2014) investigated immediate memory for hue. The present experiments required further processing of hue into long-term memory and necessitated retrieval. There are other instances of changing memory responses after time delays. The sleeper effect is a change in attitudes over time (Pratkanis, Greenwald, Leippe, & Baumgardner, 1988) with a persuasive message being followed by a discounting cue resulting in immediate low persuasion with higher persuasion over time. Pratkanis et al. suggested this is caused by two memories with differential decay combining to influence the result. Immediately the discounted message is not trusted but because the cue for the discounting information decays faster, the message becomes more trusted over time. Another example is a change in attributional judgments over time (Jacoby, Kelley, Brown, & Jasechko, 1989). Fame was attributed to names that were not the names of famous individuals after a delay. The names were initially presented on a list of non-famous names. After a day, participants did not remember the source list for the names but the names were familiar resulting in reports of the names being famous. This suggests memory for a stimulus retains its context immediately but can lose context over time. This could be decoupling of an association or it could be applying categorical criteria at response that produce different categorizations than had been applied at encoding. The context of many similar hues may be task parameters that cause distancing or atypical estimation bias. The delayed recall may result in a loss of that context with participants choosing hues that are more familiar and more typical.

4.6.4 Broader Implications

Representational shifts have been a subject of debate in the past several years because their existence would have far reaching implications for theories of categorization, memory, the influence of language on thought, and beyond. Previous work has discussed such implications from the point of view of the original representational shift hypothesis and the minimum specificity view. With the current experiments, it is clear neither previous theory accounts for our data. Below we discuss how the revealed memory response shifts, with weaker shifts due to labeling as a consequence of directed attention, might affect our understanding of other cognitive processes and phenomena.

4.6.4.1 Category Structure

Semantic memory can be conceived of as being made up of memory traces or exemplars (Heit & Barsalou, 1996; Mack, Preston, & Love, 2013; Nosofsky, 1986; Palmeri & Nosofsky, 2001). Each trace is a building block that contributes to overall knowledge of the world. Small shifts during trace creation could represent a bias in the build-up of exemplars underpinning categories, particularly if the shifts are biased by the existing category structure. The newly formed shifted representations become part of the categories that influenced their shifting. These exemplars would proceed to influence the exemplar distribution making up the category. Rather than categories being an accumulation of raw statistical experience, the distortions in new exemplars would systematically affect the content of categories. The affected category distributions would proceed to influence shifts in future experiences creating a feedback loop. Even if the bias seen in the current experiments was exclusively introduced at response, memory of the response could override the previous memory or create another as has been shown in memory reconsolidation research (Nader & Einarsson, 2010).

If new representations prove to be systematically shifted towards the prototype, the bias would be to maintain the pre-established categories through increased similarity at the expense of other possible knowledge organizations. If the representations were to shift systematically away from the prototype, as reported in Kelly and Heit (2014), the bias would undermine the current organization and support new ways of conceptualizing a semantic space. We have now shown both directions of shifts in different experimental conditions suggesting a dynamic, pervasive influence of past experience and existing categories on new representations. The context surrounding perception affects the direction of shifting and therefore the bias introduced to categories in memory.

There appears to be a dynamic influence of the environment on the interplay between perception, memory and language in real time.

4.6.4.2 Estimation Bias

As mentioned in the introduction, the Category Adjustment Model (CAM; Huttenlocher et al., 1991, 2000, 2007) appears to be consistent with the observed shifts. CAM is a model of estimating spatial location from memory which puts the introduction of bias during response. Items are encoded with some level of specificity as they are perceived. Then this exemplar information and the relevant category information combine while estimating and producing a response. The category information has more influence on less specific exemplars.

Applying a generalization of CAM to the current results, the hue exemplars given more attention would be encoded with more specificity in the bottom-up exemplar information, resulting in less ambiguity able to be affected by top-down category information. In the present results, when hues were featured at study either through labeling hues or through instructions to attend to hues, the exemplars were remembered with less bias, consistent with the bottom-up/top-down trade-off of CAM.

While the current results are broadly consistent with CAM, there are issues with applying the model directly. Specifically, CAM assumes that encodings are unbiased. Yet in Kelly and Heit (2014) we found atypical shifts near encoding. One possibility is encoding is not unbiased but can be influenced by tasks and other online cognitive processes. See Bae et al. (2015) for an application of CAM to color working memory at and near encoding.

To emphasize, we are not claiming that our results are exclusively accounted for by the CAM framework or that our results distinguish CAM from other competing models. We are seeking to illustrate that our results can be reasonably fit into an existing well-developed categorization framework. It is likely they can fit into others as well, such as a straightforward exemplar activation model, Minerva II (Hintzman, 1984), or a model that incorporates existing category information, the integration model (Heit, 1994).

4.6.4.3 Language and Thought

Knowing a language with color category distinctions leads to categorical perception effects – faster, more accurate discrimination of items crossing a category boundary than equi-different items, by some objective measure, that are within a category – while not having distinct categories in a language leads to an absence of categorical perception effects (Roberson, Hanley, & Pak, 2009; Winamer et al., 2007). This is an example of language relativity, the idea that the language(s) an individual speaks affects the way in which they conceptualize the world (Whorf, 1956).

Labels guide and facilitate category formation (Lupyan et al. 2007; Gureckis & Goldstone, 2008). In the experiments reported above, all memory is shifting relative to the category prototypes. It is not a giant leap to suggest that because English has the categories red and green, hue memory moved relative to those categories. In Greek which divides blue into two basic categories (Winamer et al., 2007), light blue and dark blue, or in Korean which divides green into two basic categories (Roberson et al., 2009), bright green and dull green, it is likely that hue memory would move relative to those different prototypes and category boundaries for speakers of those languages. This is a question that could be tested empirically.

Presently in addition to the overall shift toward the prototype and away from the boundary in line with linguistic category knowledge, there was a distinct role of overt labeling increasing memory accuracy. Online labeling appeared to direct attention to color, producing more accurate encoding. In this role of language, the things and situations a language has labels for, and perhaps habitually draws attention to, would be regularly encoded with more accurate detail than the things and situations that do not have labels.

Rather than being able to tell a simple story of language distorting reality towards the semantic knowledge encoded in the language, we have two opposing language effects. Globally, language helps guide categorization producing a memory distortion effect yet also has a local, online effect of enhancing veracity, at least in the way language was used in the present task.

4.6.5 Conclusion

We have provided novel evidence for prototypical response shifts in memory for hue. These shifts can be seen and measured in a unidimensional testing space. Existing category knowledge influences and biases recognition memory responses. Category-consistent labeling as well as other means of drawing attention to a particular dimension of an object systematically enhances the accuracy of recognition memory for that dimension. The evidence does not suggest that perceptual processing of bottom-up information is hijacked by top-down prototypical knowledge activated by labels. Nor does the evidence suggest that labeling causes shallow percepts sufficient to determine the appropriateness of the label and no more. There are pervasive influences of existing knowledge on recognition. Language plays a role in modulating those influences. It is clear that even when language is consistent with a percept at encoding, language influences how percepts are subsequently remembered.

4.7 Appendix A

The stimuli for this set of experiments were complex. More detail on the specific items and hues is reported here. Additionally, to alleviate concerns that individual hues might be eliciting different categorizations, we reanalyze encoding data from the main experiments to detect if there are any differences in judgments by hue.

Table 4.2.
Item Randomizations

Set #	Silhouette	Hue 1	Hue 2	Condition A	Condition B	Condition C	Condition D
Set 1	*Seahorse*	g1	r2	Color	Other	Other	Color
	Butterfly	g2	r3	Color	Other	Other	Color
	Giraffe	g1	r4	Color	Other	Other	Color
	Moose	g2	r2	Color	Other	Other	Color
	Seal	g2	r2	Color	Other	Other	Color
	Sunglasses	g3	r2	Color	Other	Other	Color
	Sailboat	g3	r3	Color	Other	Other	Color
	Car	g1	r4	Color	Other	Other	Color
	Watering Can	g1	r3	Color	Other	Other	Color
	Hammer	g4	r1	Color	Other	Other	Color
Set 2	*Rat*	r4	g1	Color	Color	Other	Other
	Fish	r2	g4	Color	Color	Other	Other
	Cat	r3	g3	Color	Color	Other	Other
	Parrot	r2	g3	Color	Color	Other	Other
	Bat	r4	g3	Color	Color	Other	Other
	Scissors	r3	g4	Color	Color	Other	Other
	Shirt	r1	g4	Color	Color	Other	Other
	Airplane	r3	g1	Color	Color	Other	Other
	Sword	r3	g4	Color	Color	Other	Other
	Lamp	r2	g3	Color	Color	Other	Other
Set 3	*Snail*	g3	r4	Other	Color	Color	Other
	Horse	g2	r3	Other	Color	Color	Other
	Cow	g2	r4	Other	Color	Color	Other
	Polar Bear	g2	r1	Other	Color	Color	Other
	Rabbit	g3	r1	Other	Color	Color	Other
	Genie Lamp	g4	r3	Other	Color	Color	Other
	Key	g3	r2	Other	Color	Color	Other
	Cloud	g2	r4	Other	Color	Color	Other
	Pan	g4	r1	Other	Color	Color	Other
	Badge	g4	r1	Other	Color	Color	Other
Set 4	*Camel*	r4	g4	Other	Other	Color	Color
	Penguin	r1	g1	Other	Other	Color	Color
	Kangaroo	r1	g2	Other	Other	Color	Color
	Rhino	r3	g1	Other	Other	Color	Color
	Gorilla	r4	g3	Other	Other	Color	Color
	Balloon	r2	g2	Other	Other	Color	Color
	Anchor	r1	g4	Other	Other	Color	Color
	Trophy	r1	g1	Other	Other	Color	Color
	Satellite Dish	r2	g1	Other	Other	Color	Color
	Office Chair	r4	g2	Other	Other	Color	Color

Note. Living items for the animacy condition are italicized.

4.7.1 Item Randomization

Forty silhouette shapes were created with 20 living things and 20 non-living things listed in Table 4.2. Eight stimulus lists were created rather than a full randomization. Five living things and 5 non-living things were assigned to each of 4 sets. Each silhouette shape was randomly assigned a green and a red target hue. Two of the sets were green for each

Table 4.3
Hue Coordinates in LCH and Lab Color Spaces

Hue		L*	C	H	L*	a	b
g1--		42.842	42.909	135	43	-30	30
g1-	g2--	42.842	42.909	139	43	-32	28
g1	g2-	42.842	42.909	143	43	-34	26
g1+	**g2**	42.842	42.909	147	43	-36	23
g1++	g2+	42.842	42.909	151	43	-38	21
	g2++	42.842	42.909	155	43	-39	18
Focal Green		*42.842*	*42.909*	*157.039*	*43*	*-40*	*17*
g3++		42.842	42.909	159	43	-40	15
g3+	g4++	42.842	42.909	163	43	-41	13
g3	g4+	42.842	42.909	167	43	-42	10
g3-	**g4**	42.842	42.909	171	43	-42	8
g3--	g4-	42.842	42.909	175	43	-43	4
	g4--	42.842	42.909	179	43	-43	1
r1--		41.221	79.347	14	41	77	19
r1-	r2--	41.221	79.347	18	41	75	25
r1	r2-	41.221	79.347	22	41	73	30
r1+	**r2**	41.221	79.347	26	41	71	35
r1++	r2+	41.221	79.347	30	41	69	40
	r2++	41.221	79.347	34	41	66	45
Focal Red		*41.221*	*79.347*	*36.143*	*41*	*64*	*47*
r3++		41.221	79.347	38	41	62	49
r3+	r4++	41.221	79.347	42	41	59	53
r3	r4+	41.221	79.347	46	41	55	57
r3-	**r4**	41.221	79.347	50	41	51	61
r3--	r4-	41.221	79.347	54	41	46	64
	r4--	41.221	79.347	58	41	42	67

Note. The bold hues were shown at study. The focal hues are for reference; Participants did not see the focal hues.

participant and one of the green sets would be in the color categorization condition and the other would be in the preference (Experiment 1) or animacy condition (Experiments 2 and 3). For example, reading Table 4.3, in condition 2C, Set 1 and Set 2 are red and green, respectively. They are both in the preference/animacy condition. In Experiment 4, the same lists were used with the preference condition using the 'color' lists from the chart and the animacy condition using the 'other' lists.

4.7.2 Color Calculations

The hues were calculated in L*CH color space then translated into L*ab color space for input into Adobe Photoshop. The device-independent colors were translated into RGB device-dependent color values via individual monitor color profiles created for three Dell UltraSharp U2410 monitors by a X-rite i1 Display Pro color calibrator to ensure color constancy across screens.

In L*CH color space, hue is represented by degrees of a circle. The target hues were chosen both clockwise and counter-clockwise from the focal colors reported in Sturges and Whitfield (1995). The saturation (C) and lightness (L) were held constant at the focal values.

4.7.3 Hue Accuracy Analysis

An assumption of our study is that participants could consider the displayed hues to be members of their intended category. We did a cross-experiment analysis of color naming accuracy to determine if any hues were less likely to be named accurately. Not every participant saw every hue in the naming condition – i.e., in Table 4.3, the green sets 2 and 3 are each missing one of the four green targets, a product of randomization. We limited the analysis to just those participants who saw the full range of hues in the naming condition. We also continued to exclude participants with low accuracy and/or response patterns indicating lack of engagement. Within this group, accuracy across all hues was 92.7% with individual hue accuracies ranging from 88.9% to 95.6%. Because Experiment 3 had different instructions before the naming task, which could potentially influence performance during the encoding task, we included instruction condition as a factor. In a two-way ANOVA on naming accuracy, with the 8 target hues as a within-subject factor and 2 instruction conditions as a between-subject factor, there was no difference in naming accuracy by hue, $F(3.95, 169.71)=0.71$, $p>0.05$, $\eta^2=0.02$, $BF_{10}=0.02$, using the Greenhouse-Geisser correction for violating the sphericity assumption. Based on the inverse Bayes factor, $BF_{01}=47.46$, there is very strong evidence that there are not differences by hue. There was no main effect of instruction condition nor was there an interaction effect (both $p>0.05$). These results indicate that participants were able to reliably identify less typical hues as members of the intended category as well as they could identify the more typical hues.

Similarly, there was no significant differences in proportions of 'like' responses by hue in participants seeing all hues in the preference condition in Experiments 1 and 4, $F(3.87, 116.09)=1.54$, $p>0.05$, $\eta^2=0.05$, $BF1_0=0.15$, again using the Greenhouse-Geisser correction. This result indicates that there were also no systematic biases in preference judgments for the different hues, e.g., participants did not like the "redder" reds better.

4.8 Appendix B

Two additional experiments were conducted, a control experiment accessing memory for the color category of a silhouette at test, and a norming experiment confirming that our calculated typical and atypical foils were in fact more or less typical of the color category than their target hue.

4.8.1 Category Memory Test

In the experiments reported above, participants made judgments about colored silhouettes and were tested on their hue memory in fine detail choosing among five hues spaced apart 2 degrees in L*CH color space along the circular hue dimension, approximately just noticeable differences. Each shape was presented in the original hue seen by the participant along with that target's foil hues. However, by giving participants the category information at test and forcing a response among these slightly different hues, out-of-category errors were not possible. Therefore, there is no measure of memory for the color category itself. Perhaps participants had no memory of the color of the silhouette at all and were guessing completely, with a typicality bias. To investigate color and shape bindings in our materials, we conducted a post-hoc control experiment examining color category memory for the silhouettes at test. We used the same study procedure as Experiment 2, replacing the five-alternative forced-choice recognition memory test with a cued category memory test.

4.8.1.1 Method

4.8.1.1.1 Participants

69 undergraduate students were recruited from UC Merced's participant pool using the same criteria as in Experiments 1-4.

4.8.1.1.2 Stimuli

The study portion of the experiment used the target hues as in Experiments 1-4. The test portion of the experiment used gray (L*ab hue: 82,0,0) versions of the silhouettes.

4.8.1.1.3 Procedure

Participants underwent the instruction and study procedure of Experiment 2. They made color and animacy judgments, and they had no foreknowledge of a memory test. Immediately after study, there was a memory test block. Each trial consisted of a light gray version of one of the silhouettes. Participants indicated if the silhouette had been green or red earlier with a button push. All forty silhouettes were tested, with the order randomized for each participant. The response keys were the same as the color-labeling task during study, which were counterbalanced between participants. After completing this task they continued on to the typicality task reported below.

4.8.1.2 Results and Discussion

10 participants had color categorization accuracies below 80% and/or animacy categorization accuracy below 70%. These participants were excluded resulting in a sample of 59 participants.

Overall, responses were 67% correct. There was an overall significant difference between the percent correct and chance (one sample t-test, mu=0.5, $t(58) = 12.06$, $p<0.001$, d=1.57, $BF_{10}>1000$). We used a 2 x 2 Repeated-Measures ANOVA with color (red, green) and condition (color, animacy) as within-subject factors to investigate whether these factors affect color category memory. There was a main effect of color, $F(1, 58)=5.09$, $p<0.05$, $\eta^2=0.08$, $BF_{Forward}=1.55$, with silhouettes correctly

Table 4.4.
Response Rates by Condition and Hue Category

Judgment Condition	Hue Category	Proportion Correct (Std. Dev)
Color	Green	0.62 (0.17)
	Red	0.69 (0.15)
Living	Green	0.67 (0.16)
	Red	0.68 (0.15)
Overall		0.67 (0.16)

remembered as red more often than correctly remembered as green, though the Bayesian analysis suggests the difference is only slightly more likely to occur if the effect is included in the model. There was no main effect of condition, $F(1, 58)=1.43$, $p>0.05$, $\eta^2=0.02$, $BF_{Forward}=0.23$ nor an interaction effect, $F(1, 58)=2.82$, $p>0.05$, $\eta^2=0.05$, $BF_{Forward}=1.01$. Encoding condition did not affect the memorability of the color category of a silhouette, in fact the Bayesian analysis favors the null hypothesis with a 1:0.23 or ~ 81% likelihood the condition factor will not improve the model.

4.8.2 Typicality Norming

Color display can be difficult to control. The focal, or the most prototypical, colors we used as the basis for determining all other hues and determining the direction of 'towards the prototype' were device dependent values translated into device independent space as discussed in Appendix A. This transformation along with issues of ambient lighting could result in inaccurate focal hue coordinates. We calibrated the displays and only used experiment rooms with approximately equal light intensity. But, to address concerns that our final displayed hues may not follow the typicality gradient assumed, we conducted a norming experiment. Participants were shown the hues used in Experiments 1-4 and asked to rate their typicality. We then analyzed these to determine if hues we calculated to be less typical of the category than their target hue and hues we calculated to be more typical than their target hue are in fact less and more typical.

4.8.2.1 Method

4.8.2.1.1 Participants

The same sixty-nine undergraduates who participated in the category memory test above completed this experimental task.

4.8.2.1.2 Stimuli

The 24 hues were calculated and calibrated as described in Appendix A. Rather than silhouettes, the hues were displayed on uniform squares. The hues were displayed on the same monitors and in the same experiment rooms as were used in Experiments 1-4.

4.8.2.1.3 Procedure

Each trial consisted of a hue square with the question: 'How typical is this of the color red[green]?' above the square, and the response instruction: "Please respond on a scale of 1 to 5. 1 – Very Typical, 3 – Medium Typical, 5 – Not Typical At All" below the square. Each hue was displayed twice, and the hues were displayed in blocks by color category. Therefore, there were two blocks, one red and one green, each consisting of 24 trials. The blocks were counterbalanced between participants. The 12 hues of a color category were randomized twice, with all hues being displayed once before a second randomized order was displayed in the same block. This task occurred after the category memory task and was followed by the CITY color blindness test (City University, 2002).

4.8.2.2 Results and Discussion

One participant had a response pattern indicating confusion with the task instructions. This participant was excluded along with the exclusions from the category memory task above resulting in a sample size of 58.

In order to determine relative typicality for each set of 5 hues, one target and its four foils, a difference score was computed subtracting the average typicality of the two foils we label more typical from the average typicality of the two foils we label less typical, for each set of five hues used in the memory test. Positive difference scores were consistent with our assumed typicality structure. There was an overall average difference of 0.68 which is significantly above 0, $t(57)=12.69$, $p<0.001$, $d=1.67$, $BF_{10}>1000$. Red hues had a larger mean difference of 1.04 which was different from 0, $t(57)=13.00$, $p<0.001$, $d=1.71$, $BF_{10}>1000$, while green hues had a mean difference of 0.32 which was also significantly different from 0, $t(57)=4.87$, $p<0.001$, $d=0.64$, $BF_{10}>1000$.

These results indicate that even if our displayed hues were not perfectly representative of the calculated device-independent hue, they were on an appropriate atypical to typical gradient to meet the needs of our experimental design.

Chapter 5

Event Perception and Event Segmentation Theory

5.1 Introduction

Event perception is the cognitive process of understanding sensory input changes. It is part of a wider range of event-based cognitive processes which allow cognizers to comprehend and interact with systematic changes over time. This is contrasted with processes like object perception which deals with individual items and can be considered in a context free manner to some extent. In event cognition, context is an essential component of perception. Event cognition is a central set of phenomena to be explained by any theory of cognition that takes seriously the flow of information through time. Cognitive science as a field is taking more seriously the need for incorporating context and causal dependencies as key variables when examining cognitive processes. Classic static stimuli, such as pictures of objects or single words, are special cases in natural behavior. Meaningful stability can be found not just in static objects but also at higher levels of abstraction within the dynamics of events, e.g., a constant rate of change, or a consistent action sequence ending in a consistent result. Utilizing new methods, we can meaningfully detect and measure these patterns.

A prominent theory of event perception, the Event Segmentation Theory (EST), proposes that an essential component of perceiving events is discretizing - creating segments from - the continuous flow of input to a cognizer's sensory apparatuses (Zacks, Speer, Swallow, Braver, & Reynolds, 2007). There is a growing body of empirical work investigating event segmentation and other aspects of event cognition both with behavioral methods and cognitive neuroscience methods (see Radvanskey & Zacks, 2014 for extensive review). A subset of this work will be reviewed here.

In the next section, the EST will be presented in greater detail. In the review of the theory, Section 5.2, the term representation has been used without comment in the way it has been used by Zacks and his colleagues. However, as discussed in Chapter 2, representation is a loaded term in the cognitive sciences, and is the subject of a long history of theoretical and philosophical debate. Underlying this theory are fundamental assumptions about the nature of cognition that are not settled in the wider field of cognitive science. The EST's assumptions are discussed in Section 5.3, outlining how these assumptions make it a representational model and consistent with a conceptualization version of embodied cognition. The empirical work supporting the EST will be discussed in light of the current range of perspectives on the two theoretical dimensions of representation and embodiment, questioning the assumed stances of the theory on each. Finally, this review section will conclude with an evaluation of whether the assumptions of the EST in these domains are justified by the current empirical evidence.

5.2 Event Segmentation Theory

The Event Segmentation Theory (EST) is currently the most prominent theory of event perception. It was originally proposed in 2007 and has not been substantially altered. In fact, a more expansive theory of event cognition, the Event Horizon Model (EHM) has been proposed more recently using the EST as a basic framework (Radvansky & Zacks, 2014). The EHM expands the model to make more specific predictions about the influence of segmented events on memory. Given the EST's longevity, how well it has stood up to subsequent empirical testing is of key interest for furthering the event perception research literature.

The EST is a psychological model which relates sensory inputs, working memory representations, and long-term memory representations to explain how events are understood in real time. The theory was developed to account for data from behavioral and neurological empirical studies. It has been partially implemented in a connectionist network model (Reynolds, Zacks, & Braver, 2007). Since its proposal, the EST has been the major theoretical framework through which new event perception experiments have been conceived and interpreted. Therefore, the assumptions made in the EST have become the assumptions made by many researchers conducting event perception research.

5.2.1 Definitions and Model Description

The Event Segmentation Theory is a cognitive model of perception pertaining to a particular class of things that are referred to as events. Specifically, to the EST an event is "a segment of time at a given location that is conceived by an observer to have a beginning and an end" (Zacks & Tversky, 2001). Additionally, it is concerned with "events that involve goal-directed human activity and are of modest duration (seconds to tens of minutes)" (Zacks et al., 2007). This means that events are relative to a cognizer in a particular time and space that are immediately experience-able. A two-week vacation may be described as an event in everyday language, but it is too long of a time span to be dealt with in this theory. As will be explained in detail further on, this is a theory about working memory, which has a limited duration.

Perception is defined as "a roughly hierarchical process in which sensory information is successfully transformed into representations that form the basis for action" (Zacks et al., 2007). In the model, perception is the step of processing to which sensory input is constantly supplying information. The results of perceptual processing are predictions of the "near future."

During perceptual processing, sensory input is being combined with information from the event model. In the EST, an *event model* is a working

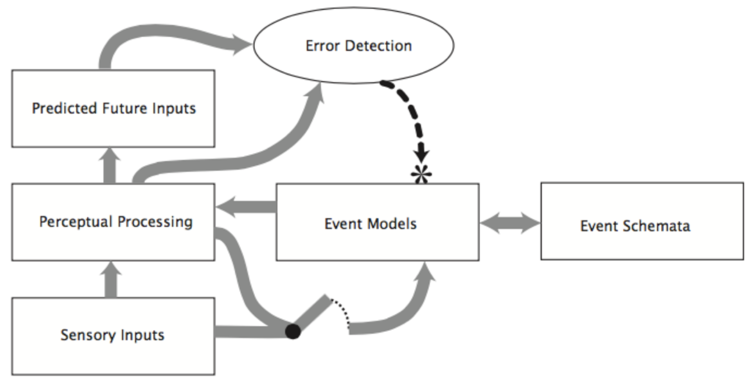

Figure 1. Schematic depiction of the theory. Thin gray arrows indicate the flow of information between processing areas. Dashed lines indicate projections that lead to the resetting of event models. The connection from sensory inputs to event models is gated, such that event models receive sensory input only during the reset phase.

Figure 5.1. Schematic depiction of the Event Segmentation Theory. Reproduced from Zacks et al. (2007, p. 274).

memory representation that is encapsulated from continuous sensory input. It is a representation of a prediction of what is expected to be happening now that is robust to transient variability of input and corresponds to current transient neural activity (separate from sensory input). The event model is formed from a combination of sensory input at the point in time the event model initiates as well as input from long-term memory representations which the EST calls *event schemata*. The event schemata capture shared features of previously encountered events which inform the sensory input to create the current event model. They are theorized to be implemented in permanent synaptic changes.

A final key concept of the EST is prediction error. As can be seen in Figure 5.1, sensory input and the current event model feed into perceptual processing. Perceptual processing creates a prediction for upcoming time points which is compared to the subsequent perceptual processing. When there is a significant mismatch between the prediction and actual sensory input, the error signal initiates a reset of the event model. This allows sensory input to feed into the working memory representation and combine with event schemata to create a new event model which will, if all goes well, reduce the prediction error. These points of high prediction error are considered to be event boundaries.

5.2.2 Key Claims

The EST claims the of gating working memory representations from continuous sensory experience results in experience of events being discretized. The segmentation of continuous experience into mixed bottom-up sensory and top-down stored-knowledge working representations is an ongoing process and is occurring at multiple timescales simultaneously. Further, these multiple timescales are hierarchically organized with shorter timescales containing discretized events within larger events. The theory claims segmentation is a mechanism of attention by conserving cognitive resources and focusing them on event boundaries. The boundaries are moments of new information that are more relevant to the individual than moments between boundaries. The EHM building on the EST framework, makes specific predictions about how the EST processes will impact long-term memory. The EST has discrete event models being stored into memory. The degree of interference or facilitation for an element of an event should therefore be relative to the number of event models in which the element. If an element is not part of the current event model, it should be less well remembered than other elements that are part of the current model. If an element is part of multiple event models that have been stored in memory, it should be more easily remembered than an element that was only in one event model that is not the currently active event model.

5.3 Assumptions

The EST/EHM takes an explicit stance that representation is exclusively internal. Radvansky and Zacks (2014) explicitly view the Event Horizon Model as amenable to embodied theories. In fact, an experiment looking at fMRI activation during reading found sensorimotor area activation corresponding to appropriate portions of the text (Speer, Reynolds, Swallow, & Zacks, 2009).

On the spectrum of embodiment stances, reviewed in Chapter 2, the strongly representational nature of the Event Horizon Model overall, and Event Segmentation Theory in particular, limits the possible embodiment stances. The EST/EHM is not compatible with the replacement hypothesis such as RECS due to the assumption of mental representation. The EST/EHM is, however, somewhat sympathetic to the constitution hypothesis. Event models are formed using a combination of perceptual input and long-term memory, making what constitution hypotheses consider external representations highly influential. However, the world does not serve as long-term memory representations in its own right. The gating mechanism cuts off external perception privileging the influence of brain-internal long-term memory influences. Given that embodiment is explicitly endorsed as relevant, event models are influenced by the environment, including the cognizer's own body, and event models form the basis of event memory, the theories are implicitly in agreement with the conceptualization hypothesis. Therefore, the EST/EHM are embodied theories but are not consistent with a paradigm shift away from traditional cognitive science assumptions.

5.4 Empirical Review

Event cognition is fundamental to the ability to interact meaningfully with the world. Many aspects of event cognition, such as predicting extended consequences beyond the immediate physical environment or understanding conventionalized gestures, appear to be representation-hungry (Clark & Toribio, 1994). At the same time, much of event cognition would seem to fit well with the concept of mutuality relations, i.e., having a change in one aspect of the system result in relational changes throughout the system. The explanation of catching a fly-ball as being a perceptual task rather than a reasoning and physics calculations task is a clear example of exploiting these relations without needing representations. To examine whether the assumptions of the EST are warranted, namely that while the body and environment may influence them, mental representations remain the privileged content of cognition, this next section will present and discuss empirical findings relying on both behavioral and neuroscience methodologies. Specifically, three claims of the EST/EHM are examined below: events are segmented into discrete temporal chunks in an automatic, ongoing fashion during perception; event processing is organized in a hierarchical format; and event segmentation structure impacts memory for events.

5.4.1 Segmentation

5.4.1.1 Segmentation: Empirical Evidence

The EST "implies that the segmentation of ongoing activity into discrete events is a spontaneous concomitant of ongoing perception and does not require conscious attention." (Zacks et al., 2007, pg. 277). The theorists suggest this interpretation is supported by imaging and behavioral data. In particular, when participants passively viewed movies of everyday tasks there is a correlation between the fMRI activation changes of that passive viewing with activation watching the same videos while actively segmenting them (Zacks et al., 2001). After viewing the movie the first time with no task, participants were asked in a second and third viewing of the movies to indicate the smallest meaningful units of activity and to indicate the largest meaningful units of activity, with one granularity of segmentation per viewing.

In the right column of the graphs within Figure 5.2, the boundaries identified by participants proceed a change in activation by ~10 seconds for both smaller and larger meaningful units. The left column shows activation change from watching the movies with no task. There were similarly timed activation changes relative to the event boundaries identified by participants during the active task. However, these activation changes were much smaller than the changes observed in the active task. The participants seem to be augmenting certain localized activation during segmentation rather than showing completely different patterns. This pattern of activation is suggested to be indicative of ongoing segmentation during passive perception.

Movement is one source of visual input thought to be causing segmentation. In order to fully specify the motion characteristics of stimuli used for segmentation, Zacks (2004) had a square and a circle move around inside a two-dimensional space. The square and circle either moved randomly as generated by an equation or were recordings of user-controlled 'intentional' shape-object movement. The movement sequences

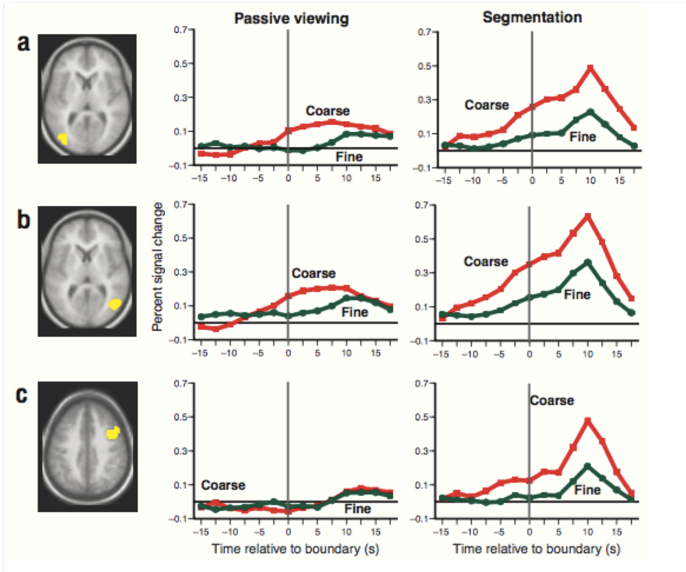

Figure 5.2. Time courses of focal brain activity in a subset of activated location. Reproduced from Zacks et al. (2001, pg. 652).

generated a range of movement characteristics such as object position, velocity, speed, acceleration, as well as secondary characteristics such as relative position and relative speed between the two objects. In Experiment 1, participants viewed the videos after a single practice video and the movement characteristics accounted for 14-16% of the variance in segmentation points. Experiment 2 gave participants more experience with how the object movement was generated. In this case, movement accounted for 26-40% of the variance in fine-grained segmentation locations and 12-24% of coarse-grained segmentation locations. Zacks, Swallow, Vettel, & McAvoy (2006) found a link between object speed and the MT+ complex but not the pSTS using the same object videos. The correlation between object speed and brain activity was stronger in the left hemisphere than in the right. No other movement characteristic showed a statistically reliable correlation with these areas. In a whole brain analysis there were a few small areas that had a reliable relationship with different movement characteristics. The MT+ complex and pSTS on both hemispheres showed increased activity following a coarse-grained segmentation, and in some cases, showed anticipatory activation before fine-grained segmentations. Motion characteristics were shown to account for a sizable portion of the variance in segmentation locations.

Indications of increased processing at event boundaries come from behavioral measures. In reading comprehension tasks, there is an increase in reading time associated with changes in key aspects of the situation model, particularly changes in time, space, causation, characters, and goals (Rink & Weber, 2003; Zwaan, Magliano, & Graesser, 1995; Zwaan, Radvansky, Hillard, & Curiel, 1998). Manipulating a clause to indicate a significant amount of time passing, e.g., ‘an hour later’ vs. ‘a moment later,’ increased the reading time of the same passage and was predictive of that clause being identified as an event boundary (Speer & Zacks, 2005). During passive movie viewing, participants' pupil diameter and saccade frequency increased at times later identified as event boundaries (Swallow & Zacks, 2004, as cited in Kurby & Zacks, 2008). These findings are used to support the EST's proposed event model flushing and rebuilding at event boundaries.

5.4.1.2 Segmentation: Discussion

Taking a contrary stance to examine the underlying assumptions about representation and embodiment, each set of evidence claimed to support segmentation presented above can be questioned. In the case of the passive viewing activation spikes, the causal relationship between relative activation levels is unclear. Some passive process is occurring which results in activation differences in these key areas as shown by the taskless viewing results. The amplification of existing activation changes with active segmentation is not necessarily an amplification of a process that is already occurring. A person does not need to discretize events in other situations to be capable of discretizing events from a continuous stream when asked to do so. When presented with the task of segmenting a video, if participants are not going to choose at random, but instead are going to have some systematicity, picking up on moments of change seems like a reasonable way to approach the task. Passively having consequences for processing as a result of changes in perceptual input is not equivalent to the elaborate EST components of constructing discrete event models. Evidence has not yet shown the passive activation increases to be exclusively attributable to the same process as the active segmentation; the passive activation increases could be due to an existing process that gets commandeered for the active task.

Similarly, motion characteristics accounting for a portion of segmentation response variance and correlating with fMRI activation changes does not decisively mean that motion changes result in discrete segmented working memory event models. When tasked with creating a response structure, relying on structure in the perceptual input stream seems more in keeping with the task than responding at random. A continuous perception and representation of events, with no moment of closing out one working memory model and building a new one, could be consistent with these results. Directly perceiving the environment, in the Gibsonian and RECS sense, would require neural changes in reaction to changes in the perceptually available environment. The evidence does not eliminate these other possibilities.

Finally with regards to the segmentation and reading results, the input during reading is inherently different from the input during other forms of visual perception. Language is a symbolic system. Letters, words, sentences, and paragraphs all have distinct, segmented structure. Finding evidence of segments in processing resulting from inherently segmented input does not provide much evidence about whether we segment continuous perception. Information is incrementally added to the existing knowledge of the narrative. At the same time, the key input changes used in these studies tend to be types of words rather than clearly segmented grammatical changes. When reading moves beyond the surface structure of text, many theories suggest situation models are created simulating the deeper meaning, but it is unclear how segmented situations models are Situation models are theorized to be rich, sensory simulations of events (Zwaan, 2004). Updating that simulation, even if it is being continuously updated, rather than discretely replaced as in the EST, requires processing and time. A slow down in gathering static information while the previous information is being processed is independently possible without the proposed working memory flushes. The processing of new information could have a heavier processing load that renewing information already active in the perceiver. Active, automatic segmentation is an unsupported leap beyond this simpler process.

Collectively, the evidence presented is consistent with automatic segmentation if you make the assumptions of the EST. However, nothing in these experiments precludes the possibilities of continuous event perception, and, if representation is theorized, continuous event representation.

5.4.2 Hierarchical Structure

5.4.2.1 Hierarchical Structure: Empirical Evidence

Event Segmentation Theory claims that perception is hierarchically organized with more fine-grained events within the larger coarse-grained events. The event model creation loop of prediction, prediction error, and model updating is theorized to be occurring at multiple time-scales simultaneously. From behavioral segmentation tasks, a hierarchy can be inferred. Comparing distributions of fine-grained event segmentation to coarse-grained event segmentation by the same individuals for the same videos, there tend to be fine-grained event boundaries just before coarse-grained event boundaries (Hard, Tversky, & Lang, 2006; Zacks, Tversky, & Iyer, 2001). Event boundaries at a finer timescale are related to the larger event boundaries identified in the coarse segmentation task. From experiments using reading, Abbott, Black, & Smith (1985) showed that reading about fine-grained tasks primed retrieval of the coarse-grained tasks of which they were part. The opposite priming pattern was not found. This is interpreted within the EST as consistent with shorter timescales more rapidly transitioning between event models with multiple of these shorter events constituting the longer timescale events.

Recent work using fMRI has been used to support the assertion of multi-timescale, hierarchical processing of ongoing events. Baldassano et al. (2017) used a data driven approach to determine event boundaries from the fMRI data of participants watching a video. The model identifies distinct brain regions processing information at different time scales. The algorithm looked for stable activity patterns which were considered to be the distinct events. Sensory areas had faster transitions between more stable patterns while higher-level brain areas had longer, fewer stable patterns. Additionally, comparing the event transitions identified by the model to event segmentation behavior in humans, there was a strong correspondence of the higher-level brain area transitions to the human identified boundaries. The lower-level areas such as V1 had many more transitions, most of which did not correspond to the human segmentation data. The higher-order patterns appear to correspond to the behavioral segmentations while the lower level perceptual areas tend to correspond to the changes within ongoing events.

5.4.2.2 Hierarchical Structure: Discussion

The hierarchical structure claim of the EST is not particularly controversial as far as representation and embodiment are concerned. There is structure in the world as well as structure in the brain. Identifying something as a mental representation or not a mental representation does not change the fact of structure existing. A mental representational view would have external structure encoded into representations with the brain network structures producing the segmentation process. A non-representational view would instead suggest that all structure has direct consequences on the unfolding processes in the brain and in the environment - there would be no distinction for internal vs. external structure. Similarly, participants reacting to the same structure at different timescales is consistent with all of the embodiment stances in addition to traditional non-embodied cognitive science. With a limited task such as viewing a static perspective and clicking a button, there are not many opportunities to distinguish between external and internal processing.

The fMRI data is more interesting for investigating these assumptions. Clearly, there are processes occurring in different brain locations for the different time-scales of processing. There is a relationship such that low-level sensory areas transition more often between stable states than later processing areas thought to coincide with comprehension of the unfolding events over longer time periods. These transitions could be distinct event representations feeding into the level above, each being tested against the current event model prediction. Or these patterns could be continuous hierarchical processing with each level being more abstract from the levels below but directly accessible to perceptual input. A major issue with using the Baldassano et al. (2017) analysis as evidence of segmented and discrete event representations with a hierarchical organization is that segmentation and discretization are assumed by the algorithm. The algorithm identifies periods of relative stability and periods of change within the neural activation patterns. To the extent that there is structure in the environment, corresponding structure in neural activation would be expected regardless of whether those neural activations are representational or are brain-based causal reactions to changes in the agent's environment.

Hierarchical organization, particularly at the neural level, appears to be factually a component of event perception. The findings from both behavioral and neuroscience experiments are largely agnostic to the theoretical concerns of representation and embodiment. The findings are certainly consistent with the EST but they do not distinguish its assumptions along these two dimensions from the other possibilities.

5.4.3 Memory

5.4.3.1 Memory: Empirical Evidence

The Event Horizon Model makes predictions of the consequences for memory if the proposed event perception architecture of the EST is true. If sensory information is only periodically allowed to affect working memory event models, and long-term memory event schemata are formed from those event models, then sensory information from event boundaries should be memorable while information between boundaries should not be as memorable. Additionally, information from before the last event boundary should be less accessible than information about the current event. The EHM specifically predicts that memory is hindered by intervening unrelated event models while it is facilitated by information being consistent across multiple event models.

Memories for items in previous events after an event boundary are diminished in comparison to memory of objects within the current event. This reduced memory has been shown in reading words (Bower & Rinck, 2001; Glenberg, Meyer, & Lindem, 1987; Morrow, Greenspan, & Bower, 1987; Morrow, Bower, & Greenspan, 1989; Rinck & Bower, 2000; Speer & Zacks, 2005), reading picture stories (Gernsbacher, 1985), and watching videos (Swallow, Zacks, & Abrams, 2009). It has also been shown for perceiver-initiated event boundaries in the case of a participant walking from one room to another, in real or virtual space, with memory for objects in the previous room being diminished in comparison to memory for objects if the participant moved within the same room (Radvansky & Copeland, 2006; Radvansky, Krawietz, & Tamplin, 2011; Radvansky, Tamplin, & Krawietz, 2010). This location updating effect has even been found when participants merely imagined moving into another space (Lawrence & Peterson, 2016). A facilitation effect has been found for the same information encountered across multiple events such as moving rooms, switching windows on a computer, and moving through different narrative contexts while reading (Pettijohn, Thompson, Tamplin, Krawietz, & Radvasky, 2016). The EHM interpretation of the results across these studies suggest past event models are less accessible than current event models since the event model is stored and rebuilt at segmentation points, but that items present across multiple event models are facilitated by being encoded more times into long term memory.

Memory for event sequences is also affected by the structure of event boundaries. When shown a video that was either (1) intact, (2) had frame deletions from an event boundary, or (3) had frame deletions between event boundaries, participants were worse at remembering the order of events in only (2) the event boundary deletion condition (Boltz, 1992; Schwan & Garsoffky, 2004). Additionally, from experiments with dementia patients compared with cognitively-normal older adults (Zacks, Speer, Vettel, & Jacoby, 2006), participants who performed a segmentation task that conformed to the segmentation patterns of knowledgeable individuals performed better on an event order memory task. These results show a deficit with improper segmentation, and are used to support the idea that event memory is made up of discrete representations and that the representations are a byproduct of regular perception.

In an attempt to separate out the introduction of new information from processing of event boundaries, Pettijohn and Radvansky (2016) introduced foreshadowing of event changes prior to the event boundaries. When an event was foreshadowed, there was no increase in reading time at the event boundary. Using memory probes, the memory patterns predicted by the EHM were unaffected by foreshadowing, which could indicate event segmentation reliably occurs even if the processing load is diminished.

Complementing the behavioral data, an fMRI study found different network activation patterns for previous event objects and current event objects that seem to coincide with long-term and working memory networks (Squire & Zola-Morgan, 1991). Further, Baldassano et al. (2017) found the identified longer timescale event pattern endings—but not the short timescale pattern endings—were predictive of increases in hippocampal activation. The strength of the hippocampal response was predictive of the strength of a similar pattern recorded when participants were recalling the events of the video from memory. The authors suggest the spike in activation is the beginning of the process of storing the previous event in long-term memory.

5.4.3.2 Memory: Discussion

Memory is clearly affected by the structure that is routinely used by participants to guide their segmentation decisions. The assumption of discrete event representations being the unit of memory is not so clearly supported by the empirical data. In the case of memory being less available after event boundaries, the boundaries have been found to be moments of change within the perceptual stream. When the information being remembered is no longer active in perception, more of the elements of an event that were concurrently perceived are likely to also be no longer active. Memory being facilitated by items crossing multiple event boundaries could be explained by having a more diverse set of concurrently perceived event elements that can serve as cues for retrieval. More integrated elements of a network are more easily accessed than more isolated elements.

Turning to the issue of representation, memory is what Clark and Toribio (1994) call a representation-hungry process - representations are an appealing explanation for how information can be stored across time in the brain. The fMRI data in particular seem to lend credence to the idea that current events are represented in working memory while previous events are represented in long-term memory. The reactivation of previously perceived information without the information currently being directly perceivable in the environment is evidence that information is stored in the brain. Coming from a perspective where mental representation is not an explanation but a placeholder (Chemero, 2009), non-representational memory storage could be a build up of activation patterns that correspond to perceptual experience that becomes stable through repeated exposure, e.g., Hebbian learning. The activation of some part of the pattern through a cue, either in the environment or from other neural activation patterns, leads to the activation of the stable collective pattern. This has a flavor of mental representation, but is non-symbolic and has quasi-stable activation patterns rather than discrete, consistent elements.

Embodiment has only been cursorily investigated in event memory. An agent walking through a door being a self-imposed event boundary suggests that physical action is able to affect event memory. Whether this is merely the body being the instigator of perceptual prediction error or a more integrated reflection of change within the agent-environment system has not been investigated. Memory being affected by the context that the information was perceived within both originally and subsequently is at the very least suggestive of conceptualization hypothesis. The extent to which information is stored in the environment or cued in synchrony with perceivable change in the environment are future avenues to explore.

5.5 Conclusion

Understanding event cognition is a central part of understanding cognitive systems. All information is perceived in a context. The interaction between that context and the information of interest changes what knowledge is internalized. The dynamic nature of event perception brings together many isolated topics that have been studied across cognitive science from object perception to reasoning to coordinated action.

The Event Segmentation Theory (and the accompanying Event Horizon Model) is the most developed theory of event perception in the cognitive science literature. EST/EHM is a squarely representational theory aligned with the conceptualization hypothesis within embodied cognition. While an impressive array of data has been collected within the EST/EHM framework testing specific predictions of the models, these core assumptions about representation and embodiment have not been directly investigated. This empirical review and discussion makes it clear that while the EST/EHM has a specific stance on these issues, the current empirical literature does not distinguish the chosen stances from the other possibilities.

In fact, some aspects of the EST are amenable to other stances on representation and embodiment. In broad strokes, the predictive processing account of cognition put forward by Clark (2013, 2015) could be seen as an embodied, less object-like representational sibling of the EST. The main incompatibility is in the EST's focus on discretization. The claim of the EST, "Event segmentation happens simultaneously on multiple timescales" (Zacks et al., 2007, pp. 277) is vague. As mentioned above the authors bound the timescales between seconds and tens of minutes, but how many simultaneous timescales are possible? If the are time scales are very narrowly spaced, the theory would approach continuity. The Clark framework, not limiting its preview to the time bounds the EST adopts, is able to apply the general idea of prediction error as the information driving change in neural activation.

Future empirical work should investigate these issues to both contribute to the specific understanding of event cognition and to contribute to the larger field-wide debates of representation and embodiment. In Chapter 2, embodiment was discussed as a potential paradigm shift. That some stances would constitute a paradigm shift makes distinguishing between these viewpoints imperative. Changing paradigms changes the questions that are relevant - if the assumed stance of the EST/EHM does not end up being supported, the empirical data being collected could end up testing hypotheses that ultimately do not advance our understanding of event cognition.

Chapter 6

Event Segmentation Decisions

6.1 Introduction

Event perception is the cognitive process of perceiving change and filtering that change for meaning. Traditionally, psychology has focused on static perception, i.e., the perception of isolated images or sounds that are tightly controlled. This focus has been a productive way to learn about perception, by controlling the stimuli and any relationships between the stimuli researchers isolate the cause of differences in behavior to those changes that were carefully orchestrated. But everyday perception is not static. Even static objects produce change for a perceiver. As a person moves they are in dynamic, relative, spatial relationships with any static objects as well as any non-static features of the environment, i.e., mutuality relations.

Some change is meaningful for a perceiver and some is not. While sitting in the grass on a sunny day, a light breeze will cause sensations across a person's skin and a leaf to tumble a few feet within their field of vision. The breeze can stop without it meaningfully changing the environment for the person. A child's laughter may turn into a shriek as they fall and scrape their knee while playing a game. This change has meaning for the caretaker who now needs to check on the severity of the injury and offer comfort. At the level of perceptual change, both situations are events. But the second situation is an event at a level that corresponds to semantic and pragmatic meaning changes, high-level events.

A predominant tack for investigating event perception in cognitive psychology has been through the lens of the Event Segmentation Theory (EST; Zacks et al., 2007) which proposes event segmentation as an ongoing passive process of taking continuous sensory input and discretizing it into working memory models. Chapter 5 outlines this theory in detail. Therefore, the present chapter will only discuss the claims of the theory that are relevant to the current empirical work. Readers are encouraged to read Chapter 5, Section 2 if they are unfamiliar with the EST. This theory is primarily concerned with meaningful, high-level events.

According to the EST, perceptual processing is actively predicting future sensory input in working memory from a combination of recent sensory experience and long-term knowledge. As predictability of input decreases and prediction error increases, the likelihood of segmentation also increases. EST argues that the process of segmentation is an automatic and necessary part of perceptual processing. In order to make sense of the information being received from a perceiver's environment, the perceiver breaks the input down into distinct units in real time.

The main evidence for this claim of the EST comes from the event segmentation task, a task in which participants indicate meaningful changes in activity within a dynamic stimulus such as a video. As discussed in Chapter 5, Section 4.1, the *behavioral task* of segmentation and the *perceptual process* of segmentation have been primarily linked through a single paradigm (Zacks et al. 2001) that is not unassailable in its argument that the behavioral task is a satisfactory proxy for the proposed perceptual process. Essentially, a BOLD signal is found corresponding to event boundaries in both passive viewing of a video and active segmenting while viewing of the same video within an fMRI scanner. A weaker signal was found during the passive viewing. The conclusion that event segmentation is the process occurring in the passive viewing is a logical leap. The active segmentation task could be picking up on a passive segmentation process as suggested by the EST. Alternatively, when actively segmenting, participants need to base decisions to segment on something in the videos, with change being an obvious choice. Processing information does not necessarily require segmentation; therefore, these results do not directly support the EST in the way that is frequently claimed in the literature.

The decisions that participants make while engaging in the event segmentation task have been glossed over in the existing reports of experiments using the task. The goal of this chapter is to highlight the decision-making process that participants engage in during the task.

6.1.1 Experiment Roadmap

In two experiments, this chapter will investigate the traditional event segmentation task behavior through continuous response measures. In Experiment 1, participants are tasked with the traditional event segmentation procedure, using a keypress to mark the end of an activity unit. Half of the participants exclusively do this task. In a second condition, the remaining participants both mark the end of units with a keypress and continuously report their current expectation of the end of the activity unit using a slider response with the computer mouse. In Experiment 2, the event segmentation task is again employed. Here, the keypress response is replaced with a slider response—the participant is asked to mark the end of activity units by bringing the slider marker from the left end of the track to the right end of the track. This response method extends the duration of a response and deviations from a ballistic rightward movement can provide insight into participants' segmentation decisions. Together the experiments are designed to explore the decision space around the event segmentation task.

6.2 Experiment 1

How do people reason about event segmentation? The event segmentation task has been a key experimental paradigm used by researchers examining event perception. Much of the evidence for the highly influential EST comes from variations on the basic task. According to the EST, segmentation is an automatic process triggered by prediction error; when unexpected incoming sensory information is perceived, a segmentation process takes place transitioning from one event model to another in working memory and storing the previous model in long-term memory. This theorized event segmentation process is an internal process. The behavior of indicating transitions between event units is a much more complex process requiring perception of transitions, the decision to respond, and the response behavior itself. Further, the event segmentation task asks participants to respond to activity units of different sizes, adding a selection process among identified transitions. This filtering requires deliberate decisions that coincide with some aspects of the video (or other stimuli that unfolds over time such as a written story). Automatic perception and deliberate decisions are very different cognitive processes yet in this literature they are often conflated.

In this experiment, there are two aims: (1) gathering new information about the relationship between event segmentation behavior and the self-reported predictability of a participant's own segmentation behavior while (2) replicating the event segmentation task to confirm the behavior observed here is similar to behavior in previous versions of the task. The measure of predictability is a continuous report of a participant's current expectation of the end of the current activity. The EST claims event segmentation occurs when there is high uncertainty. Therefore, taking the EST literally, and viewing the task as a close proxy for the perceptual process, participants should give a segmentation response when they are reporting a low expectation value. For the EST, segmentation is the result of comparing sensory input to the current event and finding high prediction error. Therefore, unexpected events need to occur for an event segmentation to take place. Participants should not anticipate segmentation points. If instead a participant predicts his or her own segmentation behavior, they should be simultaneously reporting a high expectation value when they give the segmentation response. With this response pattern, participants would be making predictions that include predictions of the ends of activity units. Predictable unit ends being behaviorally anticipated would suggest that the response behavior does not align with the EST's perceptual prediction claim.

6.2.1 Method

6.2.1.1 Participants

92 UC Merced undergraduate students participated in this experiment receiving 1 credit in the participant pool system. They were adults (18+) with normal or corrected-to-normal vision. Three participants were excluded before data analysis for not following instructions during the laboratory session.

We pre-registered a target sample size of 80 participants with a stopping rule of the target sample size plus any additional participants already scheduled. The target sample size was increased if there was an issue during data collection that indicated the data would not be usable and if a participant did not have at least one button press on more than

Figure 6.1. Experiment 1 response slider. The slider marker starts in the center and moves along the track in response to the cursor x position. As the marker is moved to the right to report higher expectation the color background gets more red.

one trial. In either case, another participant was added to the sample in the same condition as the original participant[7].

6.2.1.2 Materials

The stimuli for this experiment were videos of everyday activities. The videos are the same as were used in Experiment 2 of Zacks, Speer, Vettel, and Jacoby (2006). The videos are shot from a single camera angle in a continuous shot. The one practice video is of a man building a boat out of toy blocks. The four experimental videos were of different individuals, one per video, (1) doing laundry, (2) putting together a tent, (3) planting flowers in a flower box, and (4) washing a car. More detail on the videos is described in Zacks et al. (2006). The video display size was 600 x 400 pixels.

The slider used in the continuous response condition is displayed in Figure 6.1. The marker initiated at the mid-point of the track for each video trial. The track responds dynamically to the movement of the marker. It has a graded colored background from grey on the left to red on the right that is revealed to the left of the marker. To the right of the marker, the track has a washed-out version of the background. Revealing the more densely red portion of the slider corresponds to a higher expectation of the end of the current activity unit. The movement of the marker is tied to the x-cursor position, therefore participants only needed to move the cursor, they didn't need to select and drag the marker. The slider was centered under the video and had a width of 400 pixels.

The experiment was displayed in full screen mode on Dell Ultrasharp U2410 monitors with a 1920 x 1200 resolution.

---

[7] These criteria were introduced in a correction to the original pre-registration after data correction began. The original criteria were used to trigger exclusion and replacement of participants who did not meet the standard rather than keep trials in which they participated.

6.2.1.3 Procedure

The experimental session began with up to four participants being greeted in the main laboratory room. Participants read an informed consent form and gave verbal consent to participate. Each participant was assigned to one of four experiment rooms, which they entered, shutting the door behind them. The participant sat at a computer and read through the segmentation instructions. The sequence of the experiment from that point was a practice video, an opportunity to ask any additional questions of the experimenter, a block of two main video trials, a repeat of the instructions with a change in the requested size of the unit of activity being marked, the same practice video, another opportunity to ask questions of the experimenter, another block of two video trials with new videos, and finished with a series of open response debriefing questions.

Each participant was assigned to either the button press only (BPO) response condition or to the button press and slider (BPaS) response condition. Regardless of condition, the participants were instructed to identify “natural and meaningful units of activity” and to press the spacebar when a unit of their definition ended. In the button press and slider condition, participants additionally used the mouse to move a slider marker on a horizontal track. They were instructed to indicate if they thought the end of the current activity unit was imminent on the sliding scale. The left end of the slider track corresponded to certainty that the current activity unit would be continuing, and the right end of the track corresponded to certainty that the current activity unit would end very soon.

All participants participated in both of instruction grains, one per block. The coarse-grain instructions asked them to mark off the “LARGEST natural and meaningful units of activity” while the fine-grain instructions asked them to mark off the “SMALLEST natural and meaningful units of activity.” Each participant was randomly assigned to an instruction order. The four experimental videos were seen once by each participant and were in a random order. Therefore, all videos were watched at both grains and in both grain instruction orders across- but not within-participants.

6.2.1.4 Response Collection

A video trial began after a participant pressed the spacebar to moved forward from the previous instruction screen. The videos were set to autoplay. During the trial, measurements of the x- and y-coordinates of the cursor were collected every 50ms with a timestamp. In the condition with the slider response measure, the slider marker position was also recorded at the same 50ms intervals. The button presses and mouse clicks were also recorded with a timestamp relative to the start of the video.

6.2.2 Results

In each trial, participants either watched a video and pressed the spacebar when they decided a “natural and meaningful unit of activity” had ended, or did the watching, the key pressing, and also moved the mouse to report on a slider their current expectation of the end of an activity. Pragmatically, the slider corresponded to their expectation of needing to press the spacebar. Participants did their assigned task(s) for four videos of everyday activities, two marking the largest activities and two marking the smallest activities.

To preview the results, the analysis and results fall into two broad categories by dependent variable: discrete keypress responses and continuous slider trajectories. In the analysis of the keypress responses, participants marked off more small activities than large activities. Using group aggregates, the patterns of segmentation were not related across the fine- and coarse-grained activity sizes for the same video. However, at the trial level, most trials had substantial similarity in the location of keypress responses to other trials of the same video and instruction grain. The keypress times of a trial were more similar to other trials of the same response group - trials that only required keypress responses had more similar patterns to other keypress only trials compared to trials that collected the keypress and slider responses. Trials that collected both responses were more similar to other trials that had collected both responses. Using the continuous trajectory data, a number of movement types were evident and will be discussed qualitatively. Relating the trajectories to the keypresses, there were two predominant patterns: Either participants knew the activity end was coming, reported high levels of expectation and pressed the spacebar while the slider was in the high range; or participants pressed the spacebar when reporting a random level of expectation.

6.2.2.1 Segmentation Metrics

6.2.2.1.1 Event Duration

Variable

A key measure in the event segmentation task is the length of events identified. Average event duration is the length of the displayed video in milliseconds divided by the number of segmentation responses recorded during a trial plus one[8]. Therefore, if a participant gave one segmentation response, the average event duration would be half the length of the video.

Hypotheses

If the instruction manipulation worked, participants should segment more in the fine-grain instruction condition when they were to mark the smallest units of activity than in the coarse-grain instruction condition when they were to mark off the largest units of activity. Therefore, the hypothesis is: The average event duration is longer in coarse-grained segmentation than in fine-grained segmentation with no

[8] This was initially incorrectly pre-registered as the denominator being the number of segmentations. A correction was posted.

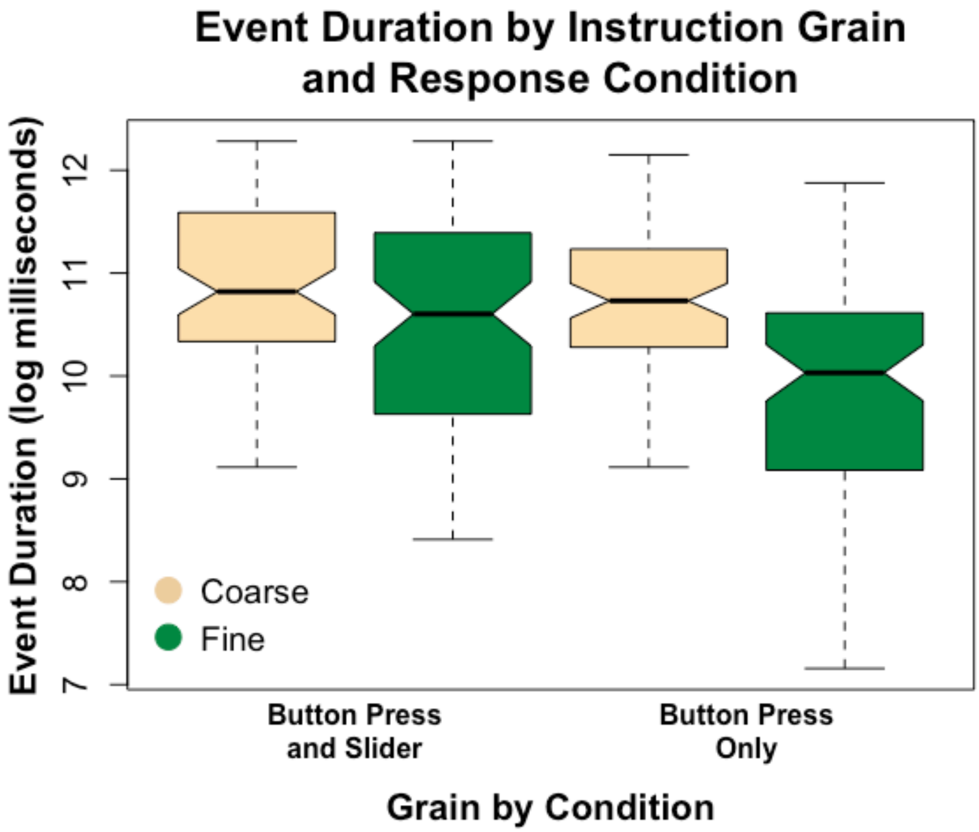

Figure 6.2. Boxplot of log event duration by instruction grain and response condition.

differences by condition. Testing this hypothesis with a linear mixed effects model, the hypothesis would be supported with a significant main effect of grain, driven by longer average durations in coarse trials than fine trials, with no interaction of grain and response condition. If a difference by response condition is observed, however, a follow-up hypothesis is: The simple effect of grain is significant in the BPO response condition. This prediction is based on the BPO condition being more similar in procedure to previous experiments in which the effect has been found.

Analysis

Average event duration was used as the dependent variable in a linear mixed effects (LME) model. The LME models throughout this chapter were run in R using the lme4 package (Bates, Mächler, Bolker, & Walker, 2014). The planned model has seven fixed effect terms. The three main effect terms are instruction grain, response condition, and instruction grain order. The four interaction effect terms are grain by condition, grain by grain order, condition by grain order, and the three-way interaction of grain by condition by grain order. The model also has two random effects: participant and video. Only the grain variable is a within-subject manipulation, and therefore, the participant effect has random intercepts

Table 6.1
Log Event Duration Linear Mixed Effect Model, Experiment 1

	Converged Model			Dropped 10 High Leverage Points		
Fixed Effects	Estimate	SE		Estimate	SE	
Intercept	**10.514**	**0.115**	***	**10.476**	**0.118**	***
Grain	**0.572**	**0.131**	**	**0.592**	**0.128**	**
Response Condition	**0.371**	**0.184**	*	0.325	0.183	.
Grain Order	-0.118	0.186		-0.186	0.188	
Grain by Response Condition	-0.469	0.221	.	-0.478	0.225	.
Grain by Grain Order	-0.234	0.219		-0.264	0.213	
Response Condition by Grain Order	-0.265	0.354		-0.385	0.355	
Grain by Response Condition By Grain Order	0.235	0.372		0.074	0.502	
Random Effects	Variance	SD		Variance	SD	
Participant (n = 82)						
Intercept	0.583	0.763		0.560	0.748	
Grain	0.525	0.724		0.529	0.728	
Video (n = 4)						
Intercept	0.022	0.148		0.025	0.157	
Grain	0.034	0.184		0.028	0.168	
Response Condition	0.012	0.109		0.012	0.109	
Grain Order	0.015	0.124		0.019	0.136	
Grain by Response Condition	0.056	0.237		0.055	0.234	
Grain by Grain Order	0.053	0.230		0.034	0.184	
Response Condition by Grain Order	0.010	0.099		0.014	0.118	
Grain by Response Condition By Grain Order	-	-		0.420	0.648	
Residual	0.172	0.414		0.161	0.401	
Number of Trials	*318*			*308*		

*** = $p<0.001$, ** = $p<0.01$, * = $p<0.05$, . = $p<0.1$

and random slopes by grain. All independent variables were within-item manipulations; therefore, the video factor has random intercepts and slopes by the full set of seven fixed effects terms.

The dataset was filtered of trials that contained no segmentation responses. Six participants were excluded that did not have at least one trial in each level of grain instruction, required for a slope estimation by grain for each participant. The resultant data set consisted of 318 trials

Table 6.2
Log Event Duration by Response Condition Linear Mixed Effect Models, Experiment 1

	Button Press Only			Button Press and Slider		
Fixed Effects	Estimate	SE		Estimate	SE	
Intercept	**10.325**	**0.137**	***	**10.704**	**0.155**	***
Grain	**0.811**	**0.197**	**	**0.349**	**0.129**	**
Grain Order	0.028	0.269		-0.245	0.244	
Grain by Grain Order	-0.348	0.320		-0.121	0.312	
Random Effects	Variance	SD		Variance	SD	
Participant	*n = 40*			*n = 42*		
Intercept	0.603	0.777		0.575	0.759	
Grain	0.578	0.761		0.481	0.694	
Video (n = 4)						
Intercept	0.011	0.103		0.036	0.190	
Grain	0.080	0.284		0.001	0.027	
Grain Order	0.031	0.176		0.000	0.010	
Grain by Grain Order	0.111	0.334		0.129	0.360	
Residual	0.154	0.392		0.181	0.426	
Number of Trials	*159*			*159*		

*** = p<0.001, ** = p<0.01

across 82 participants. The model fit to the original event duration calculation has residuals with strong heterogeneity of variance. Therefore, the dependent variable was transformed to be the log of the average event duration, reducing the heterogeneity. All fixed effect predictor variables were deviation coded, resulting in the intercept being an estimation of the grand mean. To get convergence, the three-way interaction was dropped from the video random factor. See Table 6.1 for effect estimates.

In the pre-registration, we planned to use the log-likelihood method of dropping a term from the model then comparing the fit of the new model to the original. This method is not appropriate however because the interaction term(s) take on the variance accounted for by the main effect terms when the main effect is excluded from the model. Therefore, the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) was used to test the model with all effects in place. Using the Satterthwaite method for approximating degrees of freedom, there is significant intercept, $t(12.61) = 91.551$, $p < 0.0001$, a main effect of instruction grain, $t(6.38) = 4.369$, $p = 0.0041$, and a main effect of response condition, $t(54.36) = 2.022$, $p = 0.0481$. There was a marginal interaction of grain and response condition, $t(11.41) = -2.125$, $p = 0.0562$. See Figure 6.2 for a boxplot of the average event duration by grain and condition in log milliseconds. Using the Kenward-Roger method for approximating degrees of freedom, the main effect of response condition is, in contrast, marginal, $t(47.79) = 2.002$, $p = 0.05099$. Exploring this further by removing the ten data points with the most leverage and re-running the model, the effect estimates change slightly. With the Satterthwaite method, the main effect of response condition is marginal, and the interaction of grain by response condition is significant; with the Kenward-Roger method, both effects are marginal. Further, without the high leverage points, the full model, including random slopes for the 3-way interaction by video, was able to converge. The intercept and main effect of grain are significant while the main effect of response condition and the grain by response condition interaction are both marginal. The response condition main effect and the grain by response condition interaction do not appear to be reliable. Follow-up tests of a simple effect of grain is reliable within both the BPO condition, $t(6.364) = 4.119$, $p<0.00549$ and the BPaS condition, $t(37.2) = 2.716$, $p = 0.00995$, see Table 6.2 for effect estimates.

6.2.2.1.2 Overlap

Variable

Following Newtson (1973), overlap is the amount of synchronous segmentation between a fine-grain and coarse-grain time series for the same video. In order to calculate this agreement, the duration of a video is divided into 1-second bins. Each trial is coded into a binary-binned time series. If a segmentation response (or more than one) is recorded within the bounds of a bin, the bin is coded as a segmentation point. Because videos by grain are between-subjects in the current experiment, the group binned time series are used for comparison. To get a group binned time series, all trials of a video at a grain are summed bin-wise creating a count time series. The mean count and standard deviation is calculated. Bins with a count higher than one standard deviation above the mean are coded in the group time series as group segmentation points. The count of matching segmentation bins in the fine- and coarse-grained group binned time series for the same video is the amount of observed overlap. Expected overlap is the joint probability of the two counts.

The full dataset and each subset have different group segmentation probabilities. Therefore, the observed and expected overlap were calculated separately for the full sample, the BPO condition, and the BPaS condition.

Hypotheses

The previous experiments using the event segmentation task have found more overlap than expected by chance. Therefore, the confirmatory hypothesis is: There is more overlap than expected by chance in the full sample. This would be demonstrated with a significant chi-squared goodness of fit test. If this prediction is not found, the follow-up hypothesis is: The BPO condition will have more overlap than expected by

Table 6.3
Observed and Expected Overlap by Comparison Group and Video, Experiment 1

Group	Laundry	Tent	Garden	Carwash
Full	10 (12.86)	8 (9.50)	11 (13.69)	24 (16.95)
Button Press Only	9 (9.14)	13 (16.42)	11 (9.55)	18 (15.89)
Button Press and Slider	13 (13.22)	12 (15.60)	9 (12.41)	13 (5.77)

chance. This hypothesis would be confirmed by a significant chi-squared goodness of fit test using the overlap scores calculated using only the BPO condition.

Additionally, the two conditions are expected to be equivalent in terms of segmentation response patterns. Comparing the overlap scores calculated on each response condition, this hypothesis would predict no significant effect in a chi-square test of independence.

Analysis

The observed and expected segmentation counts from the group discretized time series (see Table 6.3) were submitted to a chi-squared goodness of fit test. For each chi-squared statistic the reported p-value is based on 2000 simulations. The full data set did not have observed overlap than differed from chance, χ^2 = 4.3346, p = 0.2429. The BPO response condition alone also did not have observed overlap that differed from chance, χ^2 = 1.2134, p = 0.7596. However, the BPaS response condition alone did have a pattern of observed overlap that statistically differed from chance, χ^2 = 10.841, p = 0.01999. The pattern of observed and expected values does not, however, show a consistent pattern - one video had more observed segmentations than expected and three had less observed than expected.

The observed group segmentation counts were compared by response condition using a chi-squared test of independence. There was not a significant difference in segmentation quantity between the two response conditions, χ^2 = 1.6131, p = 0.6842.

6.2.2.1.3 Segmentation Agreement

Variable

Segmentation agreement is an adjusted correlation measure between the segmentation patterns observed in a single trial and the segmentation pattern of the group. It is computed as described in an article by Kurby and Zacks (2011, footnote 1). As described above, the binned time series is a series of one second bins equal in length to the video. Each bin is coded 1 for at least one segment during the second and 0 for no segmentations during the second. Group segmentation probabilities are computed by adding all trials of a video at an instruction grain bin-wise, i.e., calculating the group binned time series, and dividing each bin's count by the total number of trials. The point-biserial correlation (r) of a single

Table 6.4
Segmentation Agreement Linear Mixed Effect Model, Experiment 1, Full Sample Comparison

Fixed Effects	Estimate	SE	
Intercept	**0.481**	**0.019**	***
Response Condition	**-0.122**	**0.044**	*
Grain Order	-0.018	0.033	
Response Condition by Grain Order	0.013	0.066	
Random Effects	**Variance**	**SD**	
Participant (n = 88)			
Intercept	0.016	0.126	
Video (n = 4)			
Intercept	0.000	0.020	
Response Condition	0.004	0.060	
Grain Order	0.000	0.014	
Response Condition by Grain Order	0.001	0.031	
Residual	0.023	0.151	
Number of Trials (n= 328)			

*** = p<0.001, * = p<0.05

trial time series and the group segmentation probabilities is then computed. That correlation is adjusted to account for the different number of segmentations produced across trials. The minimum correlation (r_{min}) possible between a trial with the number of segmentations observed and the group probabilities is subtracted from the observed correlation. This value is divided by the range of possible correlations calculated by the maximum possible correlation (r_{max}) minus the minimum possible correlation.

$$SegAg = \frac{r - r_{min}}{r_{max} - r_{min}}$$

This adjusted correlation was computed using three group segmentation probabilities: the full sample, the BPO condition, and the BPaS condition. A trial in the BPO condition compared to the BPO group segmentation probabilities is a within-condition segmentation agreement score. The same BPO condition trial compared to the BPaS group segmentation probabilities is a between-condition segmentation agreement score.

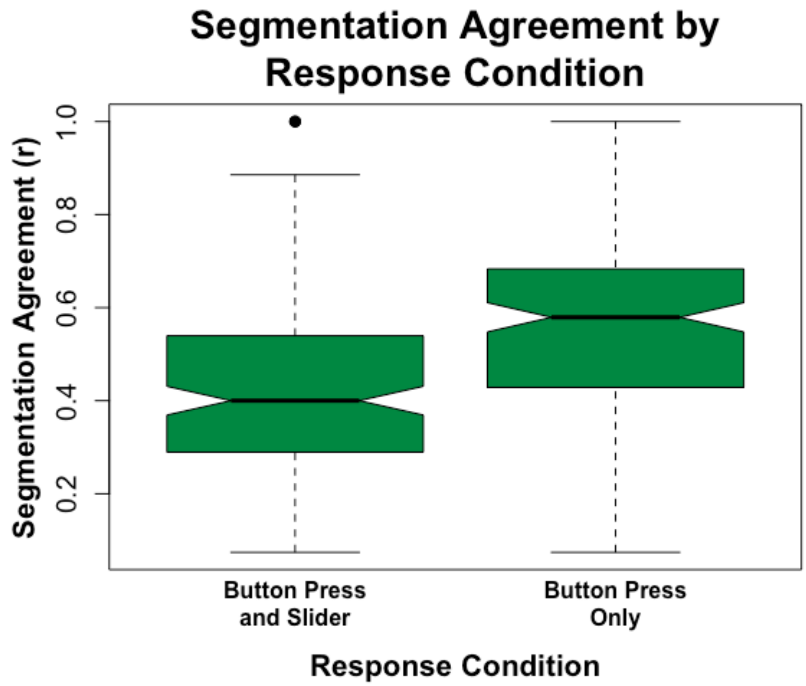

Figure 6.3. Boxplot of segmentation agreement (full sample) by response condition.

Hypotheses

Segmentation agreement is a correlational measure of how well participants agree on event boundaries and has been found in previous experiments using the event segmentation task. Therefore, the confirmatory hypothesis is: There is significant segmentation agreement between trials and the full sample, the trial's own condition, and the other condition. This hypothesis is based on the premise that the two response conditions have the same segmentation task with the keypress and will have similar behavior.

Analysis

The full sample segmentation agreement for each trial was used as the dependent variable in an LME model. The model has three fixed effect terms: the main effects of response condition and of grain order, and the interaction of the two. There are two random effects: participant and video. The participant effect has random intercepts. The video effect has random intercepts and random slopes by the full fixed effects model of three terms.

The dataset was filtered of trials that contained no segmentation responses. This resulted in 328 trials across 88 participants. The model run on the segmentation agreement scores produced residuals that have homogeneity of variance and are normally distributed. As in the event duration LME model, the fixed effect predictor variables were deviation coded, resulting in the intercept being an estimation of the grand mean.

Table 6.5
Condition Relative Segmentation Agreement Linear Mixed Effect Models Experiment 1

	Within Condition			Across Condition				Difference Scores		
Fixed Effects	Estimate	SE		Estimate	SE			Estimate	SE	
Intercept	**0.539**	**0.016**	***	**0.293**	**0.015**	***		**0.249**	**0.011**	***
Response Condition	-0.046	0.034		-0.066	0.043			0.017	0.032	
Grain Order	-0.021	0.032		-0.008	0.029			-0.013	0.021	
Response Condition by Grain Order	0.024	0.065		0.007	0.059			0.030	0.043	
Random Effects	Variance	SD		Variance	SD			Variance	SD	
Participant (n = 88)										
Intercept	0.017	0.129		0.009	0.096			0.003	0.053	
Video (n = 4)										
Intercept	0.000	0.004		0.000	0.013			0.000	0.000	
Response Condition	0.000	0.020		0.004	0.065		*BPaS*	0.001	0.034	
							BPO	0.000	0.016	
Grain Order	0.000	0.001		0.000	0.020			0.000	0.010	
Response Condition by Grain Order	0.000	0.019		0.002	0.044			0.001	0.025	
Residual	0.021	0.146		0.026	0.161			0.023	0.153	
Number of Trials (n= 328)										

*** = $p<0.001$

See Table 6.4 for effect estimates. A boxplot of the segmentation agreement scores by condition is in Figure 6.3.

Using the Satterthwaite method for approximating degrees of freedom, there is significant intercept, $t(12.27) = 25.561$, $p<0.0001$, and a main effect of response condition, $t(7.67) = -2.776$, $p=0.0251$, see Figure 6.4. There was no change in this pattern of results if degrees of freedom are estimated using the Kenward-Roger method nor is there a change if the model is re-run without the 14 points with the highest leverage in the original model.

The same LME model structure was run on the within-condition segmentation agreement scores. Using the Satterthwaite method, only the intercept was significant, $t(31.01) = 33.358$, $p<0.0001$, see Table 6.5. This did not change using the Kenward-Roger method, nor if the 2 points with extreme leverage values were excluded.

Finally, the LME model was run on the across-condition segmentation agreement scores as well. The same pattern was seen: using the Satterthwaite method, only the intercept was significant, $t(11.965) = 19.372$, $p<0.0001$, see Table 6.5. This pattern of results did not change using the Kenward-Roger method, nor if the 2 points with extreme leverage values were excluded.

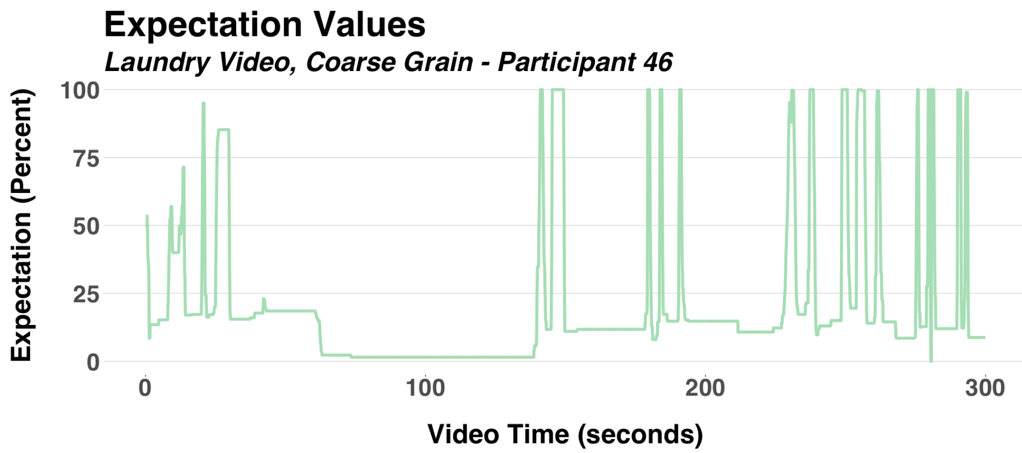

Figure 6.4. Ballistic movements in Experiment 1. The ballistic movement type is characterized by rapid changes in reported expectation.

With the significant response condition main effect in the full-sample segmentation agreement model, a follow-up examination of the meaningfulness of the effect is warranted. There was no reliable difference by response condition when using the condition specific group segmentation probabilities for calculating the correlations. However, calculating a difference score of the within-condition segmentation agreement minus the across condition segmentation agreement for each trail, it was submitted as the dependent variable in an LME model with the same effect structure[9] had a significant intercept, $t(11.678) = 22.481$, $p<0.0001$, see Table 6.5. The within condition segmentation agreement is higher than the across condition segmentation agreement. No other effect showed a difference.

6.2.2.2 Trajectories

In addition to the discrete key press responses, the forty-six participants in the BPaS condition gave a slider response with a continuous trajectory. There were a wide range of response patterns. I identified three movement types: ballistic (Figure 6.4), gradual (Figure 6.5), and stepwise (Figure 6.6). Ballistic movements are characterized by rapid changes in slider location taking one or two 50ms time intervals to move a substantial distance on the slider track, i.e., movements with high velocity. Gradual movements are movements in one direction on the track that take place over many 50ms time intervals, i.e., movements with low velocity. Stepwise movements are ballistic or gradual movements[10] that pause at a value other

[9] To get convergence, the random intercept was not correlated with the random slopes in the video random effect.
[10] The ballistic and gradual movements within a stepwise strategy were not coded as ballistic, participants needed to make separate movements of those types to be coded in those separate categories.

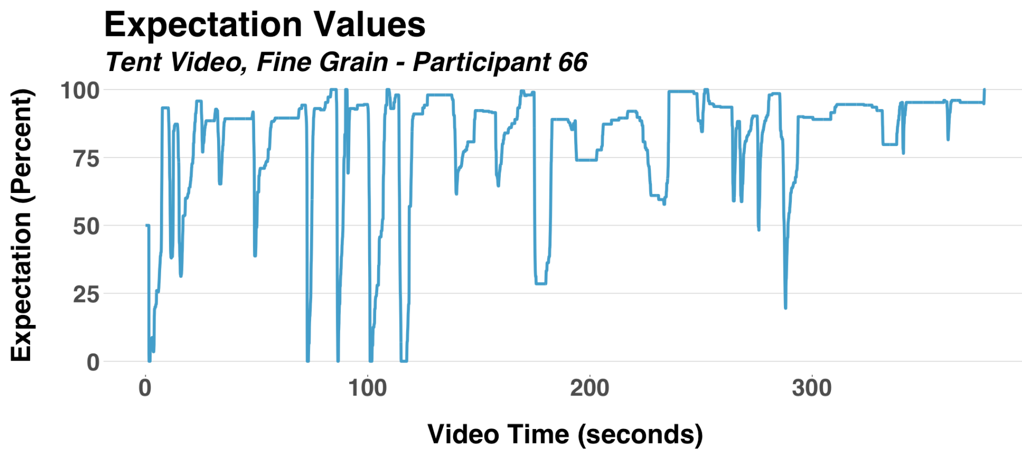

Figure 6.5. Gradual movements in Experiment 1. The gradual movement type is characterized by small adjustments in reported expectation values from one measurement to the next.

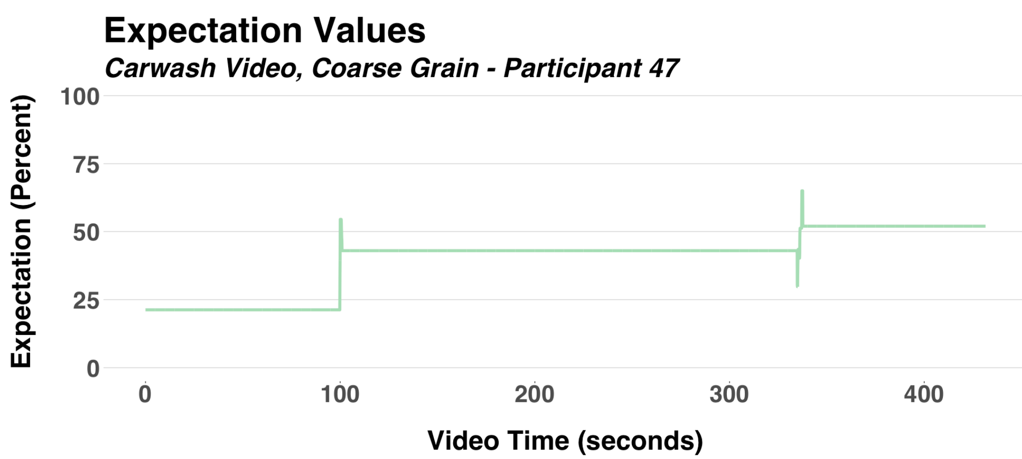

Figure 6.6. Stepwise movements in Experiment 1. The stepwise movement type is characterized by extended stopping points along the slider track. Additionally, the graphed trial which logged 6 segmentation responses is an example of the expectation response and keypress response being decoupled.

than 0 or 400. Table 6.6 contains a summary of the number of trials that fit each movement type including trials with mixed movement strategies.

Overall, in Experiment 1, participants most often used a stepwise strategy. This can be seen as deciding to move the slider higher or lower on the expectation scale and reaching a set point until they decide to adjust the expectation level again. Essentially, they chose to use the continuous slider measurement as a discrete response on a continuous scale. The second most used strategy was ballistic responses. One possible interpretation of this movement type is a participant decided to respond

Table 6.6
Response Movement Types, Experiment 1

Movement	First Block Fine	First Block Coarse	Second Block Fine	Second Block Coarse	Fine	Coarse	First Block	Second Block	All Trials
Ballistic Only	3	13	5	7	8	20	16	12	28
Gradual Only	0	4	0	3	0	7	4	3	7
Stepwise Only	16	13	22	19	38	32	29	41	70
Ballistic/Gradual	0	1	1	0	1	1	1	1	2
Ballistic/Stepwise	8	8	7	8	15	16	16	15	31
Gradual/Stepwise	5	5	4	4	9	9	10	8	18
All	4	2	1	1	5	3	6	2	8
None	6	4	10	0	16	4	10	10	20
Number of Trials Summarized	*42*	*50*	*50*	*42*	*92*	*92*	*92*	*92*	*184*
Any Ballistic	15	24	14	16	29	40	39	30	69
Any Gradual	9	12	6	8	15	20	21	14	35
Any Stepwise	33	28	34	32	67	60	61	66	127

Table 6.7
Segmentation Statistics by Slider Movement Type, Experiment 1

Movement Type	Mean Number of Responses	Range Number of Responses	Mean Event Duration (ms)	Mean SegAg *Full*	Mean SegAg *Within*	Mean SegAg *Across*	Trial Counts
Ballistic Only	6.14	(1-27)	60608.59	0.33	0.40	0.19	28
Gradual Only	6.86	(0-19)	53196.82	0.38	0.45	0.21	7
Stepwise Only	7.44	(0-56)	67758.61	0.35	0.41	0.24	70
Ballistic/Gradual	25.50	(18-33)	17703.60	0.58	0.71	0.40	2
Ballistic/Stepwise	13.06	(1-58)	49987.54	0.42	0.53	0.26	31
Gradual/Stepwise	17.78	(1-59)	47506.17	0.54	0.65	0.32	18
All	17.38	(3-39)	39746.69	0.43	0.53	0.29	8
None	5.50	(0-18)	60999.47	0.30	0.42	0.16	20

Note. SegAg = Segmentation Agreement

and waited until they moment they wanted to report higher or lower expectation levels. Alternatively, participants could be coordinating their slider movements with their decisions to respond with a keypress. Only a few trials reflect a gradual movement strategy, and even then, most gradual movement is accompanied by another movement type in the same

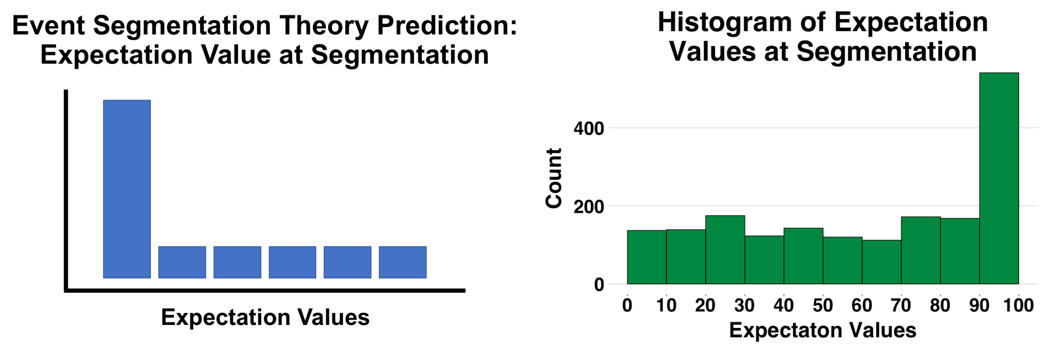

Figure 6.7. Expectation values at segmentation response. The panel on the left is the prediction of a histogram based on the Event Segmentation Theory. The panel on the right is a histogram of the Experiment 1 data.

trial. The low velocity movements could be interpreted as a participant tracking his or her estimation of the length of an activity unit, getting higher in expectation of the end of the event over time.

The different movement types had different patterns for keypresses and the derived variables based on keypresses, mean event duration and mean segmentation agreement. Participants who employed one movement strategy tended to have less segmentation responses than participants who used multiple movement types, see Table 6.7. Additionally, in spite of there being more ballistic and stepwise only trials overall, the participants who used multiple strategies had a higher correlation with the other trials across all the segmentation agreement calculations.

6.2.2.2.1 Expectation Values at Segmentation

Variable

In the BPaS condition, the participants reported a continuous expectation level of the end of the current activity unit with a slider while concurrently reporting the ends of activity units when appropriate with a keypress. The expectation values are the marker's slider track locations transformed from 0-400 pixels to a 0-100 scale. Relating the two responses to each other, the average expectation value of the measurements within a 300ms window prior to a segmentation response was calculated. This time window was selected to account for the delay between deciding to make a segmentation response and the response being made.

Hypotheses

If participants are able to predict event structure, expectation values at the time of segmentation responses should display a systematic pattern. If activity ends are unexpected, as predicted by the EST, the activity unit ends should be low and variable, see left panel of Figure 6.7 for illustration. If activity unit ends are instead predictable, expectation values at segmentation should be high. Additionally, if activity ends are predictable, grain is expected to have an influence on expectation values such that coarse-grained segment responses should correspond to higher expectation values than fine-grained expectation values. Larger activities

Table 6.8
Expectation Values at Segmentation Responses, Experiment 1, Mean (Standard Deviation)

Untrimmed	Laundry	Tent	Garden	Carwash
Fine	52.93 (29.32)	62.82 (31.51)	57.07(26.56)	69.97 (29.59)
	n=300	*n=391*	*n=363*	*n=247*
Coarse	71.32 (32.46)	53.17 (33.48)	53.93 (36.41)	63.75 (31.56)
	n=158	*n=116*	*n=194*	*n=209*
Trimmed	**Laundry**	**Tent**	**Garden**	**Carwash**
Fine	53.58 (31.47)	62.95 (31.65)	58.49 (28.92)	71.96 (30.28)
	n=258	*n=387*	*n=302*	*n=225*
Coarse	71.32 (32.46)	52.82 (34.76)	53.93 (36.41)	62.92 (32.02)
	n=158	*n=106*	*n=194*	*n=200*

should be more predictable than smaller activities. This hypothesis would be supported by a main effect of grain.

Analysis

The expectation values at segmentation were highly variable. The model has 3 fixed effects: instruction grain, instruction grain order, and the interaction. The random effect of participant has correlated random intercepts and random slopes by grain. The random effect of video has correlated random intercepts and random slopes by the full fixed effect structure.

The data were filtered of trials that did not have much movement, which is defined as different values from one measurement to the next, above .1% of trials. 7.48% of trials were trimmed with these criteria. There were 1830 remaining values across 45 participants. The means and standard deviations before and after trimming are reported in Table 6.8. As can be seen in Figure 6.7, the predominant response was high expectation values at button press, 38.74% (untrimmed: 36.4%) of responses occurred with expectation values above 80 out of 100. The remaining responses appear to occur at random points on the slider.

To improve normality, the values were squared. This transformation made the model residuals more normally distributed, but they still were not normal. Because the dependent variable is bounded, the residuals have heterogeneity of variance. The linear mixed effect model is not the ideal model for this data. However, according to Bartlett (2014), fixed effects estimates are robust to violations of heteroscedasticity though the random effects should be interpreted with caution. To get convergence, the random slopes for video by the interaction term was dropped. The estimates of the model are available in Table 6.9. Using the Satterthwaite method, the model only had a significant intercept, $t(16.80) = 9.823$, $p<0.0001$.

Table 6.9
Squared Expectation Values at Segmentation Response Linear Mixed Effects Model, Experiment 1

Fixed Effects	Estimate	SE	
Intercept	**4523.20**	**460.48**	***
Grain	349.22	495.92	
Grain Order	604.84	922.85	
Grain by Grain Order	1145.59	900.05	
Random Effects	Variance	SD	
Participant (n = 45)			
Intercept	5325132	2307.6	
Grain	5981995	2445.8	
Video (n = 4)			
Intercept	299689	547.4	
Grain	173508	416.5	
Grain Order	1213628	1101.6	
Residual	7286966	2699.4	
Number of Segmentation Responses (n= 1830)			

*** = p <0.001

6.2.3 Discussion

A number of metrics used in previous reports of event segmentation tasks were calculated and statistically tested as described above including average event duration, overlap of group segmentation time series, and segmentation agreement.

The average event duration was found to differ by instruction grain as expected. Coarse-grained events were, on average, longer than fine-grained events. The difference was larger in the button press only (BPO) condition than in the button press and slider (BPaS) condition as visualized in Figure 6.2. Average event durations were longer in the BPaS condition than in the BPO condition. The effects of response condition and response condition by grain were statistical trends so this pattern should be interpreted with caution. In light of the interaction pattern, follow-up testing was conducted, finding the difference by instruction grain was reliable in both response conditions. This pattern of results indicates both response conditions were affected by the instruction grain manipulation though it may have been more effective in the single response BPO condition than the dual response BPaS condition.

The overlap of group segmentation points across instruction grains for the same video was not greater than expected by chance, the pattern that has been found in previous research. The full sample and BPO condition did not have a pattern that differed from chance. The BPaS condition had a pattern that differed from chance, but it was not in a consistent direction. In spite of not replicating patterns seen in previous research, these results are not particularly concerning. The overlap calculated variable is quite far from the raw data, relying on group averages and arbitrary cutoffs to create the binary group time series. The videos being only seen once by each participant at a single grain makes the group comparison necessary. In Experiment 2, the videos will be seen at both grains and the within-participant version of overlap, discrete hierarchical alignment (Zacks & Tversky, 2001), is used instead which compares single trial time series to each other which is a more powerful test.

The segmentation agreement is the correlation between each trial segmentation time series and the group segmentation probabilities time series of the same instruction grain and same video, adjusted to reflect the different range of possible correlations given the number of segmentation responses of the trial. Each trial time series was compared to three segmentation probability time series: all trials, the trial's own condition (referred to as within), and the other condition (referred to as across). All three of these segmentation agreement calculations have mean correlations, i.e., the intercept estimate of the LME models, between 0.29 and 0.54 indicating that a portion of the pattern of segmentation is shared across trials of the same video and grain instruction.

The BPO response condition trials to have a higher correlation with the all-trials group segmentation probabilities than the BPaS trials correlation to the same group segmentation probabilities. Again, a possible explanation for this is the dual response required in the BPaS condition increased behavioral variability in the discrete segmentation task.

93.6% of trials correlated better with their own response condition than with the other response condition. This did not differ by response condition. It is to be expected that probabilities calculated including a trial would correlate better with that trial than probabilities not utilizing the trial. However, the mean difference from the LME model is 0.249 which is more than can be accounted for by trial's inclusion in the calculation. The conditions had somewhat different segmentation patterns.

In addition to the discrete button press response, the BPaS response condition collected continuous slider trajectories. The expectation values, which indicate the participants' expectation of an imminent end to the current activity unit as they have defined it, reflect a participant's level of expectation while the upcoming action of the video is unknown. There were a diverse set of slider movement patterns including ballistic, gradual and stepwise movement in addition to participants who did not move the slider at all. The diversity of strategies suggests participants are not all approaching the task in the same way. The diversity of trajectories may reflect different decision-making strategies. Quantitatively, the slider patterns show not all activity ends were expected or unexpected. The expectation values do not appear to have varied by grain. Over a third of segmentation responses occurred while the expectation response was high, i.e., in the top fifth of the slider track. These responses seem to reflect anticipated transitions from one event to another. The remaining values are approximately evenly spread across the rest of the range of values. In some cases, participants may have been surprised and reacting to prediction error, but many participants appeared to not correlate their expectation value responses with their keypress responses, see Figure 6.6.

6.3 Experiment 2

The standard event segmentation task features a discrete response measure - a button press. In Experiment 1, a continuous measure of predictability was elicited on top of the standard discrete response measure. In this experiment, the discrete response was replaced with a continuous segmentation response.

Responding by moving a marker along a horizontal slider track extends the previously discrete keypress response into time and space. Mouse tracking is a methodology that has been successfully employed to gain new insight into tasks that traditionally use discrete responses such as phonological competition (Spivey et al., 2005) and categorization (Dale et al., 2007). Abrams and Balota (1991; Balota & Abrams, 1995) showed that responses are not purely the product of some pre-action process. With reaction time data, researchers often discount the behavior of responding as a psychological variable. However, different cognitive variables can affect response characteristics such as force and acceleration. The patterns of response using a continuous response method for the segmentation task itself will open this additional area of variation, allowing different insight into the task. Most mouse tracking studies have two or more response options spread out on the screen allowing direction of movement to be indicative of competition between the choices. Here instead, the response is a go/no-go response where the response is either segmentation or nothing. The continuous measure allows deeper insight into the decision-making and response behavior. For example, a button press response would have no way to detect sub-threshold responses that will be evident in the slider responses.

6.3.1 Method

6.3.1.1 Participants

85 UC Merced undergraduate students participated in the experiment meeting the same criteria as in Experiment 1. They received 1 SONA credit.

7 participants were collected before an error in the experiment program was detected. The slider values were not directly being recorded. The x-cursor position and the slider value are synchronous, however, with both recording the same location along the x-dimension using a different relative point. The slider position for the affected datasets were recovered. The slider values were recorded directly for all other participants.

Figure 6.8. Experiment 2 response slider. The first panel is the starting condition with the slider marker on the left. As the slider moves toward a segmentation response on the right the darker colored bar is revealed. When the slider marker gives the segmentation response by hitting the right end of the track, the segmentation response counter increases, the slider track becomes grey and a RESET warning appears. When the slider marker returns to the left end of the track the slider resets to the starting condition.

Due to time constraints, this experiment was not pre-registered. Instead, a target sample size and stopping rule was registered after data collection was in progress. The target was to get at least 80 participants. At the time of registration, there were 29 participants already collected and up to a total of 93 would be collected. The target was to be overshot in scheduling to account for possible exclusions and missed appointments.

6.3.1.2 Materials

Three of the videos from Experiment 1 were used in the current experiment: toy block boat building (the practice video), tent building, and carwashing.

The slider for this experiment is based on the slider used in Experiment 1 but edited to encourage a respond-and-return behavior. It is visualized in Figure 6.8. The slider marker begins each trial at the left most point on the track, i.e., a value of 0. The slider has a graded background from grey on the left to green on the right, displayed in full color to the left of the marker. To the right of the marker, the background is washed out as in the Experiment 1 slider. There is a counter to the right of the slider starting at 0. When a participant responds by moving the marker all

Table 6.10

Still Frame Selection Statistics

	Boundary					
	Practice		Tent		Carwash	
Statistic	Coarse	Fine	Coarse	Fine	Coarse	Fine
Highest GSP	0.2105	0.325	0.375	0.32	0.381	0.6
Lowest GSP	0.0263*	0.1*	0.0625	0.12	0.0476	0.1
Median	0.0526	0.125	0.0625	0.12	0.0476	0.02
Count	80	99	112	87	154	107
	Non-Boundary					
	Practice		Tent		Carwash	
Statistic	Coarse	Fine	Coarse	Fine	Coarse	Fine
Lowest GSP	0	0	0	0	0	0
Highest GSP	0.0263*	0.1*	0	0.04	0	0
Median	0	0.075	0	0.04	0	0
Count	115	83	243	195	278	243

Note. GSP = Group Segmentation Probability as calculated from the Experiment 1 BPO condition. *There was an issue with the practice images that the boundary and non-boundary distributions overlapped; this did not affect the main trials.

the way to the right end of the slider track, i.e., the slider value of 400, the counter would increase by 1. A red box with white letters reading RESET would appear below the slider and the slider track would change color to grey. When the marker was brought back to the right end of the track, the reset box would disappear, and the slider track would return to its initial color scheme.

For a memory test, the fifteenth frame from every second of each video was extracted as still images. Using the button press distribution generated in the button press only condition of Experiment 1, the most segmented seconds and the least segmented seconds were identified. Specifically, using the group segmentation probabilities for each video at each grain, the 80 seconds with the highest segmentation probability were selected as well as any additional seconds with the same segmentation probability as the 80th item in the list. The same procedure was used to select at least 80 still images with the lowest segmentation probability for each video at a grain. See Table 6.10 for the quantities and probabilities. The corresponding frames to those selected seconds were used as the sample distributions from which memory test images were selected, with the highest probability images treated as event boundaries or segmentation points, and with the lowest probability images treated as non-boundaries or continuing points. Images from the appropriate distributions were randomly selected for each participant to create three pair types: boundary vs. boundary (BB), boundary vs. non-boundary (BN), and non-boundary vs. non-boundary (NN).

6.3.1.3 Procedure

The procedure was similar to Experiment 1, with only the differences reported here. Video was within-participant in this experiment. The same three videos (1 practice, 2 experimental) were used in both the fine- and coarse-grained blocks. The video order was consistent within-participant but randomized across participants. A memory test block was added after each segmentation block.

The memory test used the still images from the videos chosen from the most and least segmented seconds as described above. The number of possible images for each video at the grain of the proceeding segmentation block exceeded the number of images needed. Each trial displayed two images side by side each ~36.5% of the screen width with ~1.5% of the screen as blank space between them. The left image was labeled 1 and the right image was labeled 2, with the labels centered below its image. Participants were asked to indicate with presses of the 1 and 2 keys which of the images came earlier in the video. The trial ended after a button press or after 4 seconds. There was a 250ms blank screen between the trials. Each block consisted of 75 trials, 15 practice trials followed by 60 trials from each experimental video for a total of 150 trials across the whole experiment. Each trial set contained equal numbers of each type introduced in the materials section: BB, BN, and NN. The trials were blocked by video in the same order as the participant had seen the videos and the presentation of the three types of image pairs was randomized within each block by video.

6.3.1.4 Response Collection

The response collection scheme for the video trials reported in Experiment 1, BPaS condition, was used. The memory trials recorded button presses and reaction times.

6.3.2 Results

In each trial, participants watched a video and moved the mouse across a slider track when they decided a "natural and meaningful unit of activity" had ended. The slider marker started on the left end of the track. When they decided an activity was at its end, they moved the slider marker to the right end of the track to give the response. When the right end was reached, a counter next to the response end of the track would increase by one, the slider track would change from colored to grey and a red square with white letters would appear telling the participant to "RESET" the slider bar. The participant would need to bring the slider marker back to the left end for the slider track to become colored and the "RESET" alert to go away. The counter would not increase if the right end was reached again without first going all the way to the left. Participants marked as many or as few of these activity ends as they wished. There were two videos each seen twice, once marking small activity ends and once marking large activity ends. In between the switch in instruction grain, participants were tested on their knowledge of the order of activity in the videos they had just watched using stills from the videos[11]. They pressed the 1 or 2 key to respond. This test was repeated with different video still pairs after the second round of video segmentation.

There was one response method: the slider. First, the analysis seeks to confirm that the slider response segmentation patterns are comparable to keypress response segmentation patterns. To approximate the keypress response of Experiment 1 and of previous research, the first measurement at the response end of the slider, i.e., 400, was treated as the segmentation point. Therefore, by deriving discrete segmentation points, there are once again two types of analysis, analysis on the segmentation points and analysis of the movement trajectories. To preview the results, as expected, participants marked the end of more small activities than large activities. The segmentation pattern of individual trials had substantial similarity to the other trials of the same video marked at the same activity size. Across the two viewings of the same video, a participant was more likely to mark an activity end within the same second in both videos than would be expected if they were responding at random. By participant, the segmentation points of the two instruction grain trials were closer in time than expected by chance. For the trajectory data, participants used the same movements types as in Experiment 1. However, the new role of that slider as the only response modality required to use the slider to participate in the core task of marking off the activity unit ends. A number of strategies were employed across- and within-participants. Analyzing the trajectories statistically, the length and speed of movement towards the segmentation response within a half second of the segmentation response was not different depending on the activity size being marked or whether it was the first or second viewing of the video.

6.3.2.1 Segmentation Metrics

6.3.2.1.1 Event Duration

Variable

Instead of button presses, the segmentation response was given by bringing the slider to the left end of the track. In between responses, participants were asked to reset the slider. Therefore, segmentation times were the first measurement of the highest slider value, 400, and subsequently any first measurement of 400 after a 0 value had been

[11] An order memory test was included as a component of this experiment. Discussion of it as part of the experiment procedure is of course relevant to understanding the context of the segmentation task for the participant. However, the inferential analysis of the memory results is not reported here.

recorded. Otherwise, the average event duration was calculated as in Experiment 1.

Hypotheses

The hypothesis is again that there are longer average event durations in the coarse-grain instruction condition than in the fine-grained instruction condition. As in Experiment 1, this is a manipulation check. If participants followed instructions there should be more responses when they were asked to identify smaller events.

Analysis

Average event duration was used as the dependent variable in a LME model. The model has three fixed effect terms: two main effect terms of instruction grain and block, and the interaction term of the two. The model also has two random effects: participant and video. Both the grain and block variables are within-subject variables; however, they co- vary with a single participant seeing one grain in each block. Therefore, the participant effect has random intercepts and random slopes by grain. The variables were fully crossed by item, therefore the video effect has random intercepts and slopes by the full set of three fixed effects terms.

The dataset was filtered of trials that contained no segmentation responses. Subsequently, one participant was excluded that did not have at least one trial in each of the grain instruction levels, which is required for a slope estimation by grain for a participant. The resultant data set consisted of 333 trials across 84 participants. As in Experiment 1, the model fit to the original event duration calculation has residuals with strong heterogeneity of variance. Therefore, the dependent variable was transformed to be the log of the average event duration. All fixed effect predictor variables were deviation coded, resulting in the intercept being an estimation of the grand mean. See Table 6.11 for the effect estimates.

Using the Satterthwaite method for approximating degrees of freedom, there is significant intercept, $t(36.02) = 127.251$, $p < 0.0001$ and a significant main effect of instruction grain, $t(39.87) = 6.646$, $p < 0.0001$. The main effect of block was marginal, $t(4.41) = 2.181$, $p = 0.0882$. Using the Kenward-Roger method for approximating degrees of freedom, the degrees of freedom and some of the t-values are estimated to be a bit lower resulting in p-values are all slightly higher. Using this method, the main effect of block is also marginal, $t(4.27) = 2.181$, $p = 0.0903$. Exploring this further, removing the three data points with the most leverage and re-running the model, the effect estimate decreases resulting in lower t-values by both methods and higher p-values. The marginal effect of block appears to depend on those high leverage points.

6.3.2.1.2 Segmentation Agreement

Variable

Segmentation agreement was calculated as in Exp. 1 using the full sample.

Table 6.11

Event Duration and Segmentation Agreement Linear Mixed Effects Models, Experiment 2

	Event Duration			Segmentation Agreement		
Fixed Effects	Estimate	SE		Estimate	SE	
Intercept	**10.548**	**0.083**	***	**0.462**	**0.015**	***
Grain	**0.626**	**0.094**	***	-0.035	0.018	
Block	0.256	0.118	.	-0.042	0.024	
Grain by Block	-0.292	0.353		-0.005	0.074	
Random Effects	Variance	SD		Variance	SD	
Participant (n = 84)						
Intercept	0.496	0.704		0.006	0.079	
Grain	0.605	0.778		0.007	0.084	
Video (n = 2)						
Intercept	0.001	0.036		0.000	0.015	
Grain	0.001	0.027		0.000	0.012	
Block	0.011	0.103		0.001	0.025	
Grain by Block	0.050	0.225		0.007	0.085	
Residual	0.108	0.328		0.014	0.118	
Number of Trials	*333*			*329*		

*** = p<0.001, . = p<0.1

Hypothesis

As suggested in previous research, participants who are watching the same videos and share an understanding of the action in the video are likely to segment in a similar pattern leading to high segmentation agreement. This would be evidenced by a significant intercept.

Analysis

The segmentation agreement scores were the dependent variable in an LME model. The model had the same structure as the Experiment 2 event duration model: three fixed effects—grain, block, and grain by block—and two random effects—participant and video. The participant effect had random intercepts and random slopes by grain. The video effect had random intercepts and random slopes by all three fixed effects.

The residuals were normally distributed. There was some heterogeneity of variance with more negative residual values for the lower fitted values and more positive residual values for higher fitted values. No tested transformation fixed this issue, however, as mentioned previously Bartlett (2014) shows that fixed effects are robust to this assumption violation. Using the Satterthwaite method, the model had a significant intercept, $t(12.054) = 40.506$, $p<0.0001$ and no other significant fixed effects, see Table 6.11 for effect estimates. This did not change using

Table 6.12
Hierarchical Alignment Linear Mixed Effects Models, Experiment 2

	Discrete			Continuous		
Fixed Effects	Estimate	SE		Estimate	SE	
Intercept	**1.410**	**0.163**	***	**-2937.834**	**1057.429**	*
Grain Order	-0.463	0.406		-342.125	2093.599	
Random Effects	Variance	SD		Variance	SD	
Participant	*n=84*			*n= 75*		
Intercept	1.351	1.162		41840670	6468	
Video (n = 2)						
Intercept	0.002	0.046		111305	334	
Grain Order	0.124	0.353		266170	516	
Residual	1.556	1.247		68425410	8272	
Number of Trials	*165*			*139*		

Note. Continuous scores were trimmed to within 1 standard deviation of the mean.
*** = p<0.001, * = p<0.05

the Kenward-Roger method and excluding the two most extreme leverage points did not change this pattern of results.

6.3.2.1.3 Discrete Hierarchical Alignment

Variable

Discrete hierarchical alignment is analogous to overlap as described in Experiment 1 but within-participant instead of at the group level. A discrete hierarchical alignment score is the count of overlapped 1-second bins that contain segmentation responses in the two trials of the same video for the same participant at the two grain instruction levels minus the expected overlap calculated by joint probability. This measure was introduced in Zacks and Tversky (2001).

Hypotheses

If fine-grained segmentation responses are indicative of subsets of the activities being marked in the coarse-grained segmentation responses, there should be a higher level of alignment than expected by chance. Previous research has found higher than expected alignment. With the present model, a hierarchical alignment effect would be demonstrated by a significant intercept.

Analysis

The hierarchical alignment scores were the dependent variable in a LME model. The model had one fixed effect of grain order and two random effects of participant and video. The participant effect has random intercepts. The video effect has random intercepts and random slopes by grain order.

The dataset was filtered of missing values (due to one or both of the pair of trails not having any segmentation responses). The model fit to the hierarchical alignment scores produced residuals with a strong heterogeneity of variance. A log transformation of the hierarchical alignment scores was used instead which lessened the variance and increased normality in the residuals but these assumptions remain a concern. Only the fixed effect estimates are robust to this issue (Bartlett, 2014). Grain order was deviation coded. The effect estimates are in Table 6.12.

Using the Satterthwaite method, there was a significant intercept, $t(24.321) = 8.633$, $p < 0.0001$. Neither using the Kenward-Roger method nor dropping the highest leverage points changed this conclusion.

6.3.2.1.4 Continuous Hierarchical Alignment

Variable

Continuous alignment is the observed average distance between the nearest cross-grain segmentation points. The observed average distance is taking the average of the absolute time differences between each coarse segmentation response and its nearest fine segmentation response in the matched time series of a participant for a video at a grain. The expected average distance is calculated by computing the difference from a set of times equal to the number of coarse segmentations randomly sampled from the uniform distribution to the observed fine segmentation responses. This measure was introduced in Zacks & Tversky (2001).

Hypotheses

Previous research has found an effect of continuous hierarchical alignment. If the same participant is marking related events at the two grains, there should be a smaller average distance than expected by chance. This would predict a negative intercept.

Analysis

The continuous hierarchical alignment scores were the dependent variable in an LME model. The model has the same structure as in the discrete hierarchical alignment analysis. The continuous alignment scores are extremely leptokurtic. The residuals have heterogeneity of variance and are not normally distributed. A transformation that would make a linear model appropriate was not identified. Only the fixed effect estimates are robust to these issues (Bartlett, 2014).

Because of the leptokurtic shape, no significant effects were found using the continuous hierarchical alignment scores trimmed by three standard deviations or even two standard deviations. In an exploratory analysis, to get a picture of the pattern within the most central data points,

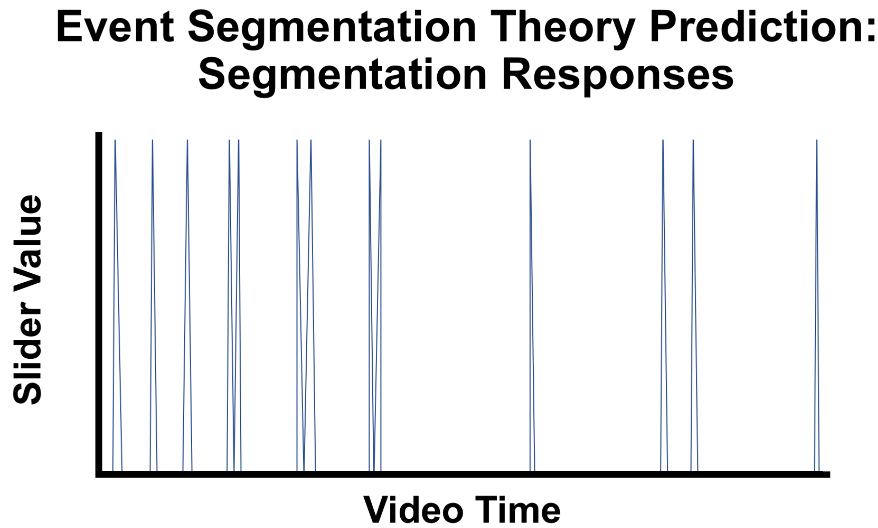

Figure 6.9. Event Segmentation Theory prediction of Experiment 2 segmentation responses. If segmentation is the reflection of a discrete perceptual process of segmentation, responses should be exclusively ballistic.

the alignment values were trimmed to scores within one standard deviation of the mean. This model had a significant intercept in the negative direction, $t(6.818) = -2.778$, $p = 0.0281$, see Table 6.12 for effect estimates.

6.3.2.2 Trajectories

A key advantage to using mouse movement responses rather than keypresses is the response movement takes place over space and time. In this task, the response movement could occur in a number of ways while moving between the reset location and the response location on opposite ends of the slider track. The only restriction on movement is the requirement to reset the slider after a segmentation response. The EST prediction for these trajectories is exclusively ballistic movements, as segmentation behavior would be the result of discrete event model changes, see Figure 6.9 for an illustration. Participants did engage in ballistic movements however a range of movements were observed.

As was seen in Experiment 1, there are three main categories of response movements. The movement could be ballistic (Figure 6.10) going straight from the left track end to the right track end. Alternatively, the response movement could be gradual (Figure 6.11) with the slider being slowly moved toward the response end of the slider. Participants also used a stepwise movement (Figure 6.12) by moving toward the response end of the slider then pausing before continuing rightward. While Figures 6.10-6.12 were chosen to illustrate the categories of movements, participants often employed two or all three movement types during a trial (see Figure

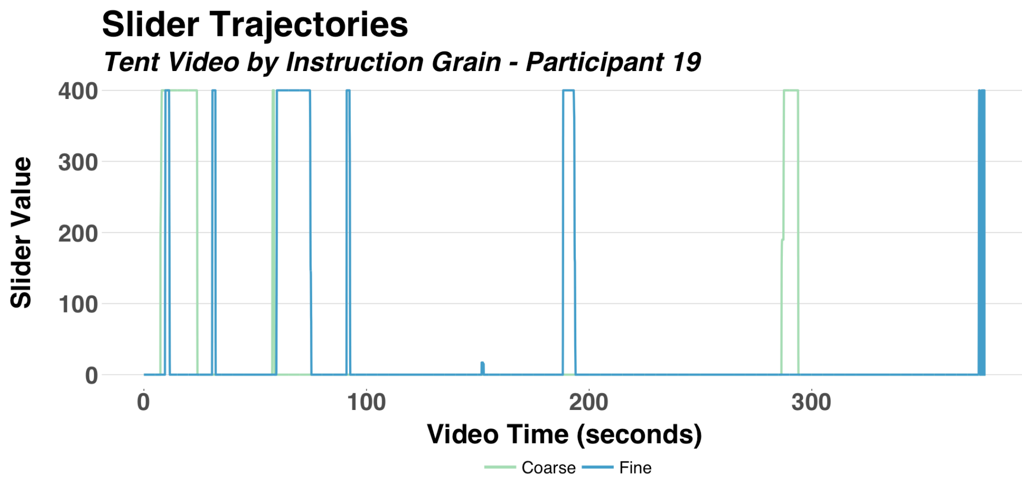

Figure 6.10. Ballistic movements in Experiment 2. The ballistic movement is characterized as participants rapidly moved the slider from the reset location to the response location.

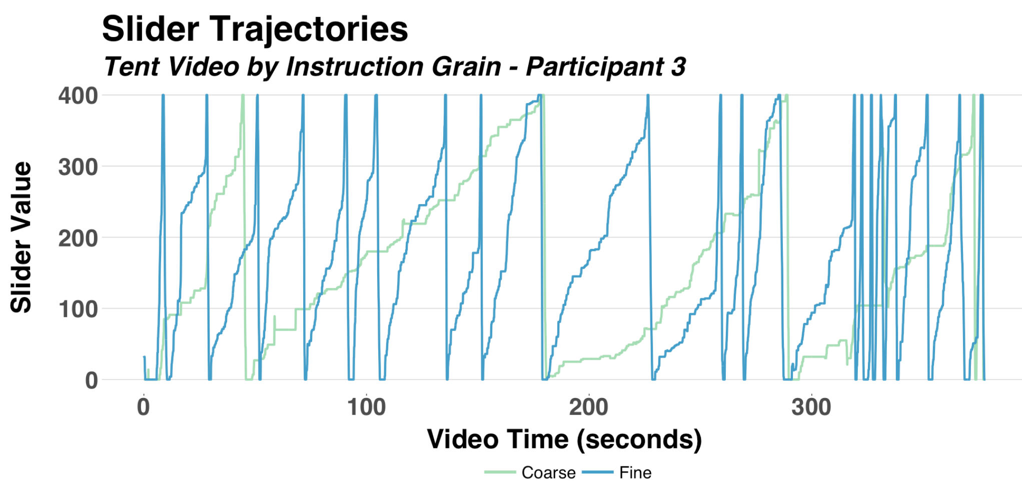

Figure 6.11. Gradual movements in Experiment 2. The gradual movement type is characterized by slow increases of slider values over time. This participant also exhibits ballistic movements towards the end of the fine-grained trial.

6.13 for an example.) The frequencies of each of trials with these movement types are reported in Table 6.13. Characteristics of the trials are broken down by movement types in Table 6.14 Investigating the EST prediction of exclusively ballistic movements, 47.1% of trials fit this type. In total, over 85% of trials had at least some discrete responses. Even though Experiment 1 demonstrated that participants can report their expectation of activity ends, participants often used the response slider as a discrete response. However, many response trajectories did take place over time and showed expectation of the coming activity unit end, behavior that does not reflect a perceptual process.

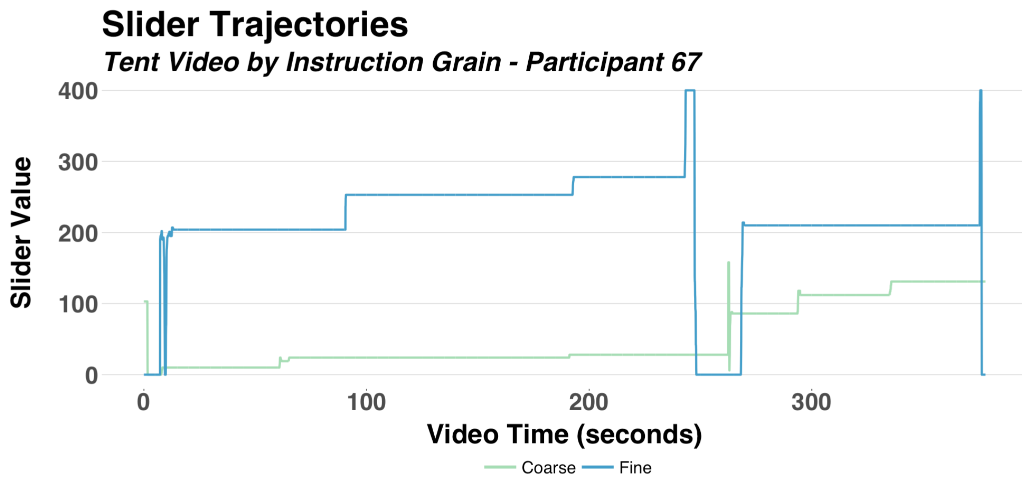

Figure 6.12. Stepwise movements in Experiment 2. The stepwise movement type is characterized by extended stopping points along the slider during the trajectory towards the response. The graphed trials also exhibit some ballistic movements. The coarse trial had stepwise movement while having no segmentation response, i.e., reaching a slider value of 400.

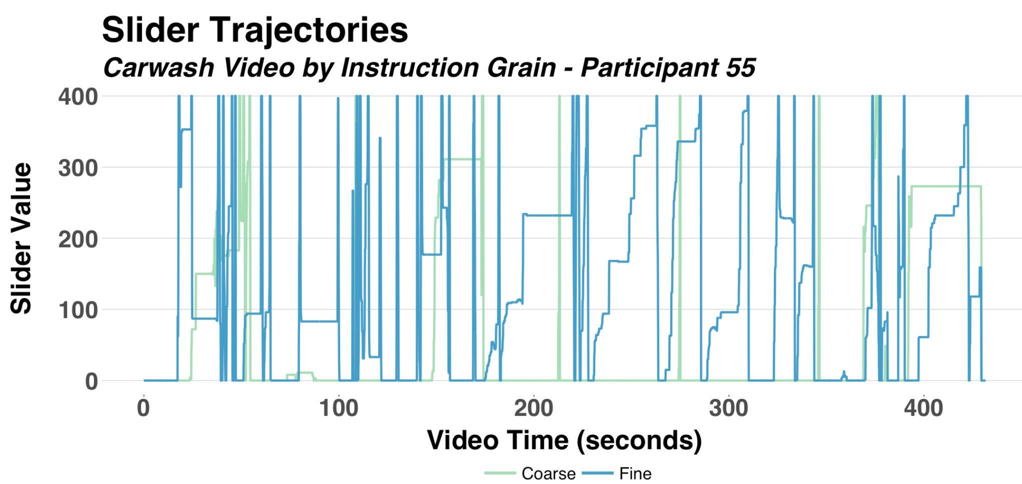

Figure 6.13. Mixed response movement types in Experiment 2. All three movement types are exhibited in both the fine-grained and coarse-grained trials for this participant.

A number of adaptive slider response strategies can be identified. Some participants used the ballistic movements of moving rapidly toward the response end of the slider track, e.g., both trials in Figure 6.10, as well as some responses in Figure 6.13, Figure 6.14, and Figure 6.15. The gradual movement was employed in a strategy of moving quickly to reset the slider then slowly moving the slider marker towards the response without specific intent to respond within milliseconds, e.g., both trials in Figure 6.11, as well as some responses in Figure 6.13, and one coarse trial response at the end of the video in Figure 6.15. Participants also used

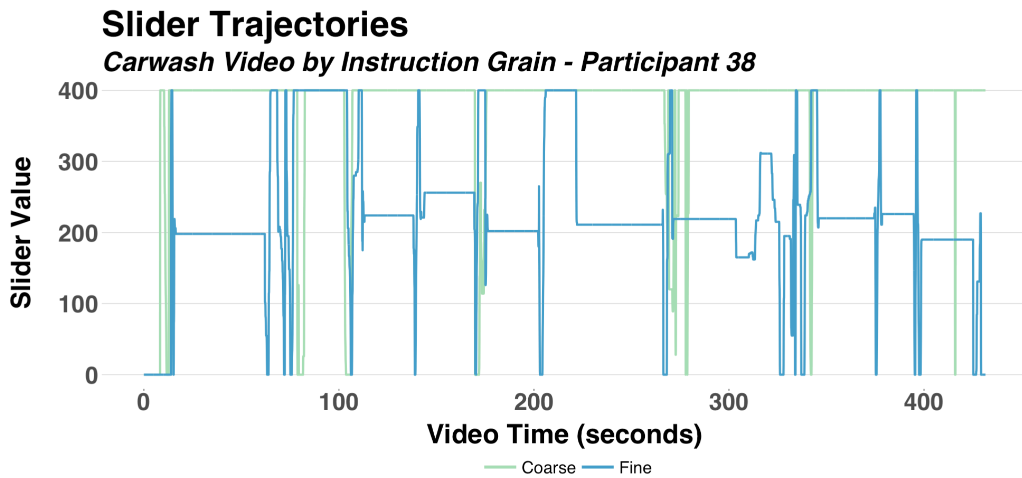

Figure 6.14. Return to the Center and Reset to Respond Strategies. The coarse trial in light green is ballistic with the resting state being the response end of the slider. The fine trial in blue uses a return to the center strategy where after ballistically responding, the participant resets the slider by going to 0 then brings the slider marker back to the middle of the track, decreasing the distance to respond ballistically.

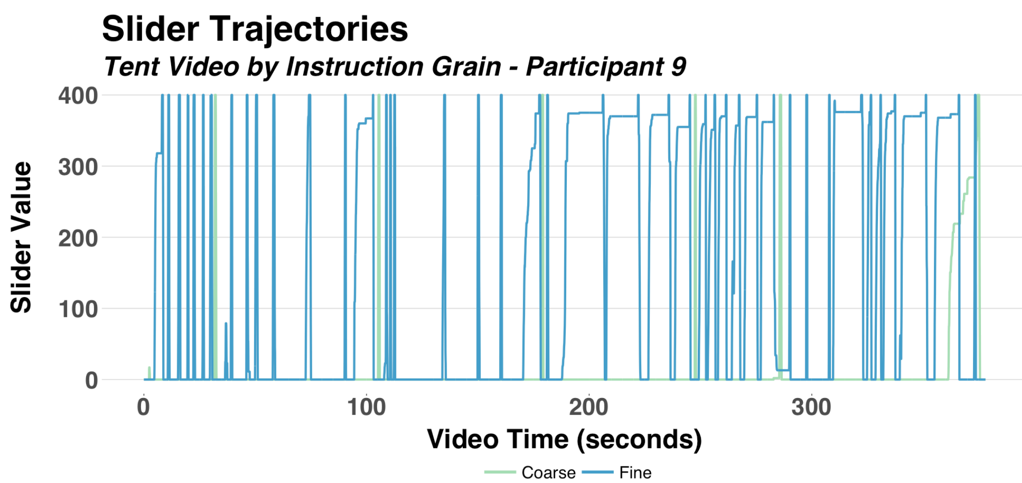

Figure 6.15. Return to Response End Strategy. The fine trial in blue exhibits a strategy of resetting the slider by moving the marker to 0 then returning to the response end of the slider track. This is a variation of the return to center strategy.

various adaptive strategies to be able to respond more quickly than they would be able to if the response started at the reset end of the slider track. In Figure 6.12, the participant moved closer to the response end in short bursts then ballistically responded over a short distance. In Figure 6.14, the fine trial exhibits a strategy of ballistically resetting the slider after a response then immediately returning to the center of the slider, waiting for the next activity end with a shorter response distance. In Figure 6.15,

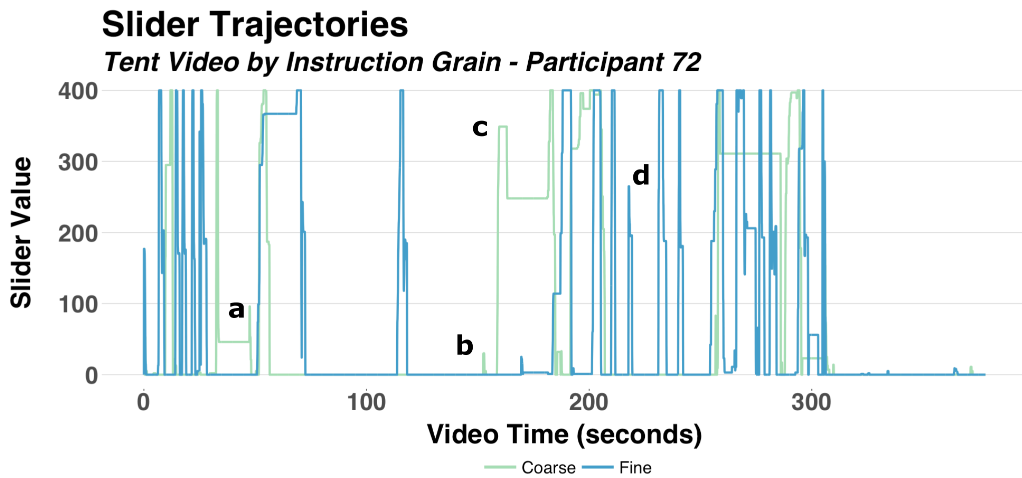

Figure 6.16. Sub-Threshold Responses. Many movements toward the response end of the slider track are not completed in these trials. From left to right, the indicated responses are examples of: (a) resetting the slider after moving toward the response end when slider is not reset, (b) small movements of a few pixels that are backtracked, (c) a seemingly "stepwise" response that is partially backtracked, and (d) a movement more than halfway across the slider that is ballistically backtracked to the reset end of the slider, though no reset is required.

the fine trial exhibits this strategy more dramatically, with the participant responding then returning the marker to within 50 pixels of the slider response, making the distance traveled from active response movement initiation and the response location much shorter than even the center strategy.

In Figure 6.14, a curious strategy is employed in the coarse trial. The participant does not reset the slider immediately after responding. Instead, he or she moved the slider marker twice the distance of the slider to reset and respond ballistically. 37 or ~10.1% of trials used this strategy for part, if not all of the video. This is surprising as the strategy appears to be maladaptive.

Another feature of the trajectory data is sub-threshold responses. Participants often would start a response movement then abandon it, backtracking toward the reset end of the slider track. In fact, ~62.9% of trials have some identifiable sub-threshold responding. Identification of these non-responses is a bit complicated. For example, stepwise movements toward the response then stopping may reflect a deliberate strategy or could reflect sub-threshold responding without backtracking. Additionally, because the slider needs to be reset, there are also movements that look like ballistic sub-threshold responses but are in actuality a belated reset movement rather than backtracking due to an aborted response decision. Each of these sub-response movements are displayed in Figure 6.16.

Table 6.13
Response Movement Types, Experiment 2

Movement	First Block Fine	First Block Coarse	Second Block Fine	Second Block Coarse	Fine	Coarse	First Block	Second Block	All Trials
Ballistic Only	44	32	38	46	82	78	76	84	160
Gradual Only	2	5	2	5	4	10	7	7	14
Stepwise Only	5	8	2	0	7	8	13	2	15
Ballistic/Gradual	10	10	10	4	20	14	20	14	34
Ballistic/Stepwise	16	14	15	17	31	31	30	32	62
Gradual/Stepwise	3	8	2	3	5	11	11	5	16
All	6	7	15	10	21	17	13	25	38
None	0	0	0	1	0	1	0	1	1
Number of Trials	*86*	*84*	*84*	*86*	*170*	*170*	*170*	*170*	*340*
Summarized									
Any Ballistic	76	63	78	77	154	140	139	155	294
Any Gradual	21	30	29	22	50	52	51	51	102
Any Stepwise	30	37	34	30	64	67	67	64	131

Table 6.14
Segmentation Statistics by Slider Movement Type, Experiment 2

Movement	Mean Number of Responses	Range of Number of Responses	Mean Event Duration (ms)	Mean Segmentation Agreement
Ballistic Only	15.66	(1-105)	49649.29	0.48
Gradual Only	14.43	(1-69)	101658.48	0.45
Stepwise Only	1.07	(0-2)	145718.41	0.29
Ballistic/Gradual	21.00	(1-77)	36692.85	0.53
Ballistic/Stepwise	11.29	(0-53)	54860.64	0.43
Gradual/Stepwise	4.38	(1-16)	116714.94	0.41
All	19.58	(3-55)	30421.93	0.49
None	0	(0)	0	0

6.3.2.2.1 Area Under the Curve

Variable

The dimensions of space and time can be used to construct a two-dimensional shape of the mouse trajectory. The area between this trajectory and an axis can be calculated for comparison. The units of the space dimension are pixels on the slider track and are bounded between 0

Table 6.15
Area Under the Curve Linear Mixed Effects Model, Experiment 2

Fixed Effects		Estimate	SE
Intercept		-1.4554	1.5066
Grain		-3.8603	2.2474
Block		0.1335	2.3625
Grain by Block		1.7995	8.6643
Random Effects		Variance	SD
Participant (n = 85)			
Intercept		29.210	5.405
Grain	Coarse	177.600	13.330
	Fine	13.840	0.410
Video (n = 2)			
Intercept		0.000	0.000
Grain	Coarse	0.173	0.416
	Fine	1.539	1.240
Block		3.805	1.951
Grain by Block		80.230	8.957
Residual		592.300	24.340
Number of Segmentation Responses (n= 4985)			

and 400. The time dimension is characterized by measurement intervals of 50ms. There are a number of ways to define a trajectory to a segmentation response, the first slider value of 400 after a slider value of 0 has been recorded. One option would be to define it as the last recorded 0 slider position measurement time before the response measurement to the response measurement time. Another option would be to pre-define a window of interest such as 10 measurements or the 500ms preceding the slider response measurement. In order to cut down on variability, a 10-measurement window was applied here. Two additional manipulations were applied to the trajectories to isolate the signal. Not all participants kept the slider at zero between responses. To account for this difference, the lowest value within the 10 measurements was deducted from all values in the sequence. The transformed sequence captures only the distance traveled by the slider marker over the 500ms window of interest. Additionally, some 10 measurement sequences captured a descent in slider values before the rise to the segmentation response. Only the ascending values captures the to-response trajectory. Therefore, rather than a blanket 10 measurement sequence of interest, we used the last zero value within the 10 measurements as the start of the trajectory. For ease of interpretation, the slider value was transformed with a division by four, creating a scale of 0-100. The final trajectory sequence was used to calculate area under the curve (AUC) - the space between the trajectory line

on its rise from zero to its peak (400 or 400 minus the lowest value). The AUCs were calculated using Simpson's rule.

Hypotheses

One possible pattern within the AUC of the final ascent within the 500ms prior to a segmentation response is a more ballistic responses, a lower AUC, could be observed for responses in the second block where the videos were being watched for a second time.

Analysis

The AUC values were the dependent variable submitted to an LME model. The planned model has the same basic structure as in the event duration and segmentation agreement models, with the fixed effects of grain, block, and grain by block, as well as the random effects of participant and video. The participant effect has correlated random intercepts and random slopes by grain. The video effect has correlated random intercepts and random slopes by the full fixed effect structure.

To improve normality and decrease heterogeneity of variance, the AUC values underwent a Box Cox transformation with a lambda of 0.4. The two data points beyond 3 standard deviations of the mean were dropped. The planned model failed to converge. To get a model that did converge, first the Box Cox transformed AUC values were centered. Additionally, the random intercepts and random slopes for both participant and video effects were uncorrelated. These changes were determined using the protocol suggested by Barr, Levy, Scheepers, and Tily (2013).

The fitted model had residuals that were normally distributed but has some heterogeneity of variance. Therefore, only the fixed effects should be considered unbiased (Bartlett, 2014). Using the Satterthwaite method, no effects were statistically significant, see Table 6.15 for effect estimates. There was a lot of variability in the amount of leverage attributed to individual data points in the model. A cluster of 29 high leverage trials were dropped from the model. The same model did not converge. Dropping the random intercept by video, the simplified model was run on the filtered data set and produced the same pattern of results.

6.3.3 Discussion

Segmentation Metrics

As in Experiment 1, a number of metrics reported in previous event segmentation research were calculated and statistically tested. Here, these metrics are average event duration, discrete hierarchical alignment, continuous hierarchical alignment, and segmentation agreement. Of key interest is whether a similar pattern of results to previous response patterns were found with the slider response method. Average event duration was found to differ by instruction condition as expected. Coarse-grained events were longer on average than fine-grained events. This shows that the grain manipulation was effective. The mean segmentation agreement adjusted correlation was 0.47, comparable to the mean agreement in Experiment 1. This suggests that there is a systematic pattern to segmentation responses for each video at an instruction grain. The mean discrete hierarchical alignment was found to be reliably above zero indicating that there was more overlap of segmentation across grains than would be expected by chance. The mean continuous hierarchical alignment was not reliably different from zero but had a distribution very inappropriate for a linear model. The alignment score skews negative which is the predicted direction of lower average distance than chance. Using a central subset of the data, it appears this trend may remain a pattern when accounting for the random effects of participant and video. Overall, with the exception of continuous hierarchical alignment, the pattern of event segmentation metric results in the current experiment are comparable to metrics observed in discrete segmentation tasks.

Trajectories

The qualitative analysis of the slider trajectories revealed a number of response patterns. There are three identified movement types. A number of different strategies, both adaptive and maladaptive, were evident across trails. These response patterns are consistent with diverse decision-making processes rather than a direct reflection of an automatic process. Finally, sub-threshold responses were observed frequently with at least one sub-threshold response in close to two thirds of trials. These sub-responses are particularly revealing; they demonstrate participants are initiating movement then deciding not to respond. This filtering process is another layer of separation between perceptual processing and responses in the event segmentation task.

The slider trajectories varied widely in temporal length and velocity across the slider. The area under the curve (AUC) of the slider response in space over time was calculated as a way to systematically look at the slider trajectories. Limited the length of the trajectories of interest to the ascending slider values within 10 measurements of a segmentation response, there was no difference by instruction grain or block. At least by this metric, there was not a systematic difference in slider movement by these groupings. Future analyses of the Experiment 2 trajectory data are discussed in section 6.4.2.2.

6.4 General Discussion

This chapter has, in two experiments, investigated the event segmentation task using a continuous response measure. A discrete response such as a keypress provides minimal data for interpretation. Alternatively, continuous responses give researchers more information about how participants are responding in a task.

Experiment 1 implemented the standard discrete event segmentation task. Videos were used as the event stream and segmentation responses were indicated with keypresses. The control condition, the button press only response condition, was a replication of the event segmentation portion of Experiment 2 in Zacks et al. (2006) including the instruction wording and video stimuli. The novel manipulation is the addition of a second condition, the button press and slider condition, with a continuous slider response reporting ongoing expectation of the end of an activity in addition to the standard keypress response. For the most part, with the exception of overlap, the metrics used to characterize event segmentation had patterns consistent with previous event segmentation task experiments. In the button press and slider condition, there was evidence of similar segmentation patterns. The differences between the response conditions indicate the addition of the second response may have increased variability in performance on the discrete response task.

The qualitative analysis of the expectation value trajectories has revealed diverse response styles. Across all trials participants have moved the slider quickly and slowly. After moved the slider marker to a new location, sometimes the slider is left in that spot and sometimes the slider is quickly moved back in the opposite direction. The diversity in response styles makes generalization over the group difficult. The first view on the data, isolating expectation values at segmentation points, reveals some systematicity and a high degree of variability concurrently. One possible interpretation is that some segmentation points are highly expected while others are a surprise. Alternatively, the random expectation values at response could reflect a lack of engagement in the slider response task.

Experiment 2 implemented the event segmentation task using a continuous response measure for the segmentation responses. The experiment also introduced a memory order accuracy test. The metrics used to characterize event segmentation were found in patterns consistent with previous research.

The qualitative analysis of the response trajectories revealed the same movement types as were seen in Experiment 1. Experiment 2 issued participants a different task with the slider, however, and therefore, while general movement types aligned the inference of the purpose of the movements is different. A number of strategies were identified: using 0 as the base-point and responding via a ballistic movement to the response end of the slider track and a quick return; resetting the slider quickly then slowly moving towards the response end, picking up speed when the end of an activity is immanent; using alternative base-points than 0 including the approximate midpoint of the track, within 50 pixels of the response, and the response location itself. Participants also used a stepwise strategy where they moved the slider toward the response end in bursts. Beyond this diversity of strategies, the slider response also revealed a large quantity of sub-threshold responses, information that is not available using a keypress response. In a first statistical analysis on these trajectories, analysis of the area under the curve of the ascending measurements within half a second of the response did not vary systematically relative to instruction grain. The area also did not differ relative to viewing order of the same video by a participant.

6.4.1 Trajectory Response Implications and Limitations

Together, these experiments demonstrate that continuous response measures are compatible with the event segmentation task. The continuous response on top of the discrete segmentation response in Experiment 1, as well as the continuous response as the segmentation response in Experiment 2 do not radically alter the event segmentation task metrics. The trajectories produced in both continuous response types are complex with a wide range of behaviors exhibited by participants.

Expectation Value Slider Responses

A continuous expectation response is a novel measure for event segmentation. The ability of an individual to predict the end of an activity is evident in the distribution of expectation values at segmentation. Over a one third of keypress segmentation responses were simultaneously reported as highly expected on the slider. The majority of the 7.48% of trials that were trimmed had mean expectation values segmentation at or near 50, the starting position. The .1% rate of change threshold was selected as a low bar for inclusion that excluded the expectation values with no movement or a single movement, however, examining the trajectories qualitatively relative to the threshold could improve the selection of trials with no response from those with a few stepwise movements. Antidotally, this condition confused the most participants, with many asking questions after they completed the practice trial and some expressing in the debriefing question they didn't know what to do with the slider. The presented LME model does not account for the movement strategies of trials, or more specifically, portions of trials that the segmentation responses fall within. The under-analysis of the relationship between segmentation and expectation values in this task is the primary limitation.

Slider Segmentation Responses

The slider segmentation responses revealed a plethora of movement types and response strategies. There was not a consistent manner of response across- or even within-participants. The various response strategies as well as the sub-threshold responses, i.e., evidence of filtering, demonstrate that the event segmentation task is a reasoning and decision-making task.

Using a continuous response with dynamic stimuli is a messy endeavor. With most mouse tracking studies, the trajectory begins at a time-locked location and the response target is available on the screen. The slider responses of Experiment 2 are analogous to these studies in that it represents a discrete response of marking a boundary while the response behavior is recorded as it unfolds over time. Unlike previous studies, with present decisions relative to a stimulus that is also unfolding over time, time-locking the beginning of the trajectory is not feasible.

Resetting after a response is also a challenge. For example, taking control of the cursor from participants would be disorienting. Yet for the response to be given by moving the cursor across the screen, the cursor position is must be brought back to that initial position to distinguish the beginning of a response trajectory. The employed technique of requiring the participant to reset the slider themselves seemed to work reasonably well. Yet resetting takes attention and time while the video plays on. With event segmentation, pausing the video would introduce new events to the participant's experience of the video.

As a first attempt at this sort of response measure with this type of stimuli the paradigm was successful reproducing most of the basic event segmentation response patterns. The quantitative trajectory analysis of area under the curve was not revealing of systematic patterns. However, as with the trajectories in Experiment 1, utilizing and improving on the qualitatively identified movement types is a promising path forward. Each segmentation response can be coded for the behavior it resulted from. I am optimistic that further exploratory analysis of the trajectory data, taking into account sub-populations of response strategies, will uncover systematic patterns.

6.4.2 Future Analyses

The initial analysis has largely demonstrated the viability of the continuous response slider in the event segmentation task. While the variability in trajectories is indicative of the main theoretical point that the event segmentation task is not a perceptual task and is not a good proxy for a perceptual process, the initial explorations of the trajectory data with pre-planned inferential analyses did not yield much specific information about what participants are doing during the task. The data sets are complex and there are many avenues of analysis that have not been explored. I will outline a few of the many possible analyses that are beyond the scope of this chapter but will be conducted in the future.

6.4.2.1 Experiment 1

The analysis in Experiment 1 could be improved via incremental changes to existing models as well as integrating the qualitative analysis of trajectory movements into quantitative models of the data.

Segmentation Agreement

An improvement to the reported LME models of within-condition segmentation agreement and between-condition segmentation agreement would be to replace the grain order fixed effect with block and grain, as was used in Experiment 2. Segmentation agreement is calculated at the video by grain level. It is possible that there are systematic differences by grain within these scores that are not dependent on the grain order.

Additionally, the descriptive statistics of the movement types revealed a pattern such that trial trajectories that exhibited one movement type tended to have lower segmentation agreement compared to those which contained multiple types. It is possible this is confounded by keypress response quantity. Both variables in the same statistical model would help clarify whether the descriptive pattern is meaningful.

Expectation Values at Segmentation

The expectation values at segmentation do not differ by the predictor variables of the LME model, however, there are a number of additional predictors that should be explored for this dependent variable. One prediction is segmentation responses that agree with other individual's segmentation responses are more expected than the responses that are more idiosyncratic. The group segmentation probability time series captures the group agreement on where event boundaries are located within the video. The group segmentation probability of a particular 1-second bin that a segmentation response falls within may predict expectation values at segmentation.

Integrating the qualitative analysis of slider trajectories into the quantitative analysis is a future analysis goal. There are three identified categories of response movements. Coding each time series for periods of movement of each type is the next step to learn more about the relationship between segmentation and expectation values in this task. For example, I predict that ballistic movements are most correlated with high expectation values. This would reflect participants syncing their two response modalities.

6.4.2.2 Experiment 2

The analysis of Experiment 2 has been focused on discrete response points, i.e., event duration, segmentation agreement, and hierarchical alignment; focused on trajectories but inconclusive, i.e., area under the curve; or qualitative, i.e., trajectory movement types and response strategies. Future analysis will seek to use inferential statistics on the trajectory data and find additional ways to group the trajectory responses. Additionally, the memory task will be analyzed.

Trajectory Strategies

There are a number of exploratory hypotheses that could be tested using the trajectory movement and strategies coding scheme. For example, because there is a higher density of responding, fine-grained trials could have more ballistic responses than coarse-grained trials. However, fine-grained trials, having more time pressure, could have higher rates of anticipatory strategies, e.g., returning to center and gradually moving towards response beginning at the reset timepoint. The anticipatory strategies reduce the distance between the location of the marker before the acceleration that culminates in a response. With more time between responses in the coarse-grained trails, a delayed response does not affect the next response as frequently.

I would also predict that participants using anticipatory strategies would have more probable segmentation points compared to the group segmentation probabilities of the same videos with the exclusive keypress response, i.e., the Experiment 1 BPO condition. The shortened response delays due to shorter distances to travel would allow the segmentation points to more closely align. This analysis presumes a lag of 0. Cross-Recurrence Quantification Analysis (cRQA) is a method of examining lag between two time series (Coco & Dale, 2014). Utilizing cRQA to find the optimal alignment between the group segmentation probabilities and each trial, I predict that anticipatory strategies would have the shortest lag, ballistic movements from the resent point would have a longer lag, and the responses that initiate at the response point traveling to the reset point and back would have the longest lag.

Sub-threshold Responses

The sub-threshold response patterns have not yet been analyzed. I predict that sub-threshold responses will occur at or just after more probable segmentation points. Participants who do not respond at a point when many participants do respond may be considering making a response. If a response is aborted because the participant did not move quickly enough, the sub-threshold response should occur on a delay relative to a high probability segmentation point. Alternatively, a response could be aborted because the participant decided the activity end was not at the appropriate size for the current instruction grain. If so, the prediction would be sub-threshold responses in the coarse-grained instruction condition correlate with overt responses in the fine-grained instruction condition.

Event-Related Trajectories

Additionally, I would like to do an analysis of the event-locked trajectories. Inspired by event-related potentials (ERPs) in EEG data, these event-related trajectories (ERTs) will be windows of the slider measurement stream aligned by the segmentation responses. This is a programmatic way of finding more nuanced movement types. For instance, while many segmentation responses are gradual, attempting to aggregate over gradual segmentation responses may reveal systematic variation leading to sub-categories that are qualitatively distinct as strategies.

Order Memory

Finally, the memory data from Experiment 2 has gone unexplored in this chapter. The motivation for conducting the order memory test is to expand upon previous findings that memory for stills from a video is greater when the stills are from event boundaries. The memory task was to choose which of a pair of stills occurred earlier in the video. The pairs have different time differences between them. Future analysis will focus on the memory accuracy as the dependent variable and include a number of predictor variables including the type of pair, e.g., boundary vs. boundary, the time difference between the pair, the quantity of participant's own segmentation responses for the seconds represented by the images of the pair, and the group segmentation probability difference between the pair.

6.4.3 Theoretical Contributions

The two experiments in this chapter demonstrate that the event segmentation task is not an acceptable proxy for an automatic perceptual segmentation process. The basic event segmentation task requires participants to respond to the ends of activity units of a particular order of magnitude, i.e., smallest or largest natural and meaningful units of the participants' own judgment. The task demands not only require participants to respond to continuous perception, but also, to decide what qualifies as an activity unit at a particular size. Rather than respond to every activity end they perceive, participants need to filter their responses. The frequency of sub-threshold responses observed in Experiment 2 confirm the filtering process is inhibiting potential responses.

The EST claims the automatic perceptual segmentation process initiates when errors in perceptual predictions pass a threshold (Zacks et al., 2007). Because using the same videos across instruction grains allows for within-participant analyses, which have greater statistical power than between-participant analyses, event segmentation experiment designs often repeat stimuli across instruction grains. As discussed in Chapter 5, participants having perceptual prediction errors aligned across instruction grains while experiencing the same video twice in the same experimental session is logically flawed. If participants have memory of the exact video during the second viewing of that video, and segmentation behavior reflects perceptual prediction error, the response rate should drop dramatically for the second viewing trial. Experiment 2 had participants segment two videos twice; Block and grain order fixed effects were not significant in any of the linear mixed effects models. Block and grain order have not been reported as a significant factor in the literature. In fact, Zacks, Tversky, and Iyer (2001) argue that no effect between first and second viewing of a video is a feature in that effects are robust to familiarity. The segmentation behavior effects may be robust, but that robustness is in direct opposition to the theory of automatic perceptual prediction error triggered responses. While this critique is a theoretical contribution and the current data supports the critique, there is amble existing evidence in the literature that supports the critique as well.

Now that I have established what the event segmentation task is *not*, what cognitive processes are being reflected by the task? The task is a reasoning and decision-making task. Beyond perceiving the continuous stimulus, participants must determine criteria for what activities count as largest or smallest activity units, which they then can use to determine activity ends. For example, on a coarse-grained trial for the carwash video, a participant could decide that soaping and scrubbing a side of the car is one activity and on a fine-grained trial that participant might mark off each scrubbing motion. Pre-determining the scope of activities would allow participants to anticipate, or expect, the end of the activity. The high expectation of an activity ending reported for many segmentation responses in Experiment 1 would fit this activity-end identification strategy. However, this is just one possible strategy; as there are many response strategies displayed across participants and trials in Experiment 2, there may also be many identification strategies. The key theoretical point, however, is participants use reasoning strategies, they do not exclusively react ballistically to unexpected activity ends. In fact, ballistic responses are predicted by the EST as a signature of discrete perceptual segmentation, but these responses could also reflect failures in reasoned anticipation. For example, say a participant planned to mark off soaping and rinsing activities in the carwashing video and had expected the steps to be sequential with all soaping to occur before all rinsing. Instead, the steps are alternated on different areas of the car. The segmentation response for the first rinsing activity would be ballistic because the participant reasoned that the soaping activity would continue but saw activity that met the criteria for a new activity. The perceptual prediction of more soaping would be violated, but so too would the high-level reasoned expectation.

When viewing the event segmentation task at the reasoning and decision-making level, how participants perceive change is not the determinant of behavior. Event changes could be perceived discretely as the EST theorizes or event changes could be perceived continuously with qualitative differences resulting in behavior triggers. For example, a recent model of event representation for reasoning (Khemlani, Harrison, & Trafton, 2015) creates markers for the types of changes that are consequential for situation modeling (Zwaan & Radvansky, 1998). In the model, reasoning takes place over these discrete markers, but the mechanism for getting to a representation of a discrete change that serves as an event boundary does not necessarily have to be discrete itself. Being able to treat something as discrete does not require a discrete base representation. The Khemlani et al. (2015) model demonstrates how an event can be effectively discrete in a particular frame regardless of having changes within its boundaries. If asked to reason about large activity units, participants can discount small activity unit boundaries even though they are perceiving the changes that the boundaries rely upon.

To summarize, the present chapter provides evidence for two broad theoretical contributions. For the EST, the present experiments demonstrate that the event segmentation task does not provide conclusive evidence for the proposed perception and working memory processes. In order to test the theory, other tasks and paradigms need to be developed keeping in mind processes beyond perception and memory affect behavior. For the event segmentation behavior itself, the segmentation task is a decision-making task. In order to use event knowledge, people need to be capable of identifying relevant event elements such as time, location, people, and objects. The event segmentation literature has provided substantial evidence that changes in elements are key to understanding events. Viewed as a reflection of reasoning in addition to memory and perception, the event segmentation task and similar paradigms can productively be used to investigate how people detect and utilize meaningful experiences that unfold over time.

6.4.4 Conclusion

This chapter has presented a novel exploration of event segmentation decision-making and has implemented innovative response measures with variations on mouse tracking methodologies. The basic event segmentation task has been replicated and appears to work with the response measures I have introduced. The trajectory analysis is preliminary, but in both experiments, participants appear to have utilized the slider in diverse and systematic ways. The pattern of expectation values at segmentation suggests participants can predict their segmentation responses, which does not align with the predictions of the EST. Responding in the event segmentation task is a decision, subject to high-level reasoning as evidenced by the multiple strategies participants used in Experiment 2. The event segmentation task is not a good test of the EST mechanistic commitments but instead is a useful paradigm for exploring event reasoning.

Chapter 7

General Discussion

7.1 Summary

This dissertation has presented theoretical and empirical reviews as well as empirical investigations. Theories of representation (and non-representation) were reviewed in Chapter 2. The chapter goes on to review theories of embodiment (and non-embodiment). The two theoretical domains are discussed in light of each other, explaining how some views on the two topics are complementary while others are mutually exclusive. Chapters 3 and 4 are empirical investigations of the Representational Shift Hypothesis (Lupyan, 2008). Chapter 3 focuses on the claim that there are label-driven shifts from the true item properties towards increased category typicality at memory trace encoding of visually presented items. The reported experiments, using colors varying along the hue dimension as stimuli, provide little evidence of an effect of labeling and instead provide evidence of ubiquitous shifts in the opposite direction from category typicality. Chapter 4 makes a second attempt at finding the predicted pattern of label-driven shifts towards increased category typicality, testing color memory along the hue dimension after a delay of a few minutes. The experiments demonstrated an overall bias towards increased category typicality regardless of labeling. However, labeling did have an effect, one of decreasing the size of the typicality bias. Switching gears, Chapters 5 and 6 focused on the topic of event segmentation. The Event Segmentation Theory (Zacks et al., 2007) was reviewed in Chapter 5 along with the empirical evidence associated with the theory. Harkening back to the issues of representation and embodiment raised in Chapter 2, Chapter 5 contains a discussion of the theoretical commitments of the Event Segmentation Theory and points out some of these commitments are assumed and remain untested. Chapter 6 is an empirical investigation of the event segmentation task, the main experimental task used to investigate the proposed cognitive process of event segmentation, the automatic discretizing of continuous sensory input as a core component of perception. In the present chapter, I will attempt to contextualize the empirical findings within the theoretical domains of representation and embodiment as well as reflect on implications of my work for the field of cognitive science.

7.2 Theoretical Implications for Representation and Embodiment Debates

The theoretical issues of representation and embodiment were introduced at the beginning of the dissertation and related to each other. The empirical chapters do not address these issues in depth. Therefore, here I will explore the relationship between the reported empirical studies and these core theoretical issues in the cognitive sciences. The dissertation contains two distinct lines of empirical research on representational shifts of hue memory and event segmentation decision-making. The investigations of these phenomena will first be discussed independently, then the connections across the phenomena will be discussed.

7.2.1 Representational Shifts

The Representational Shift Hypothesis (RSH) is a representational theory, as the name implies, but it is not a proponent of traditional symbolic object-like representations. It is a dynamic representational theory where online perception is affected by existing knowledge structures. Following the Connell & Lynott (2014) taxonomy of representations being current activation and concepts being long-term knowledge, the RSH suggests that language and labeling is a way to tap into existing knowledge quickly, activating core aspects of a concept associated with a label. This allows those core concept features to be represented actively in real time. The incoming perception is actively represented as well. The RSH suggests these two sources of information, the external world through perception and the internal knowledge through labeling get co-activated and the representation is a combination of the two. As the representation gets stored into long-term memory, the memory trace is the blended product of both sources.

As for embodiment, the RSH is at the very least embodiment compatible. It is a hypothesis about brain-based mental processes, therefore it is not a replacement theory. It does not explicitly extend representation beyond the body as a constitution theory would nor does it explicitly suggest representations are affected by our bodies as a conceptualization theory would. However, in spite of not fitting neatly within one of Shapiro's embodiment categories, RSH does argue that the history of the cognizer affects online processing, with the history of associating particular words with concepts resulting in a fast connection from an experience word to those mental concepts resulting in concurrent representation of perception and prior knowledge. RSH is explicitly concerned with the agent-environment system in so far as the presence of labels in the current environment affects the perception of other aspects of the environment.

The experiments reported in Chapters 3 and 4 do not fully agree with the RSH. The experiments of Chapter 3 tested whether an influence of labels on hue memory would be evident after half a second or after five seconds. There was no reliable influence of labels. In fact, regardless of labeling, the bias was in the atypical direction. Expanding the time delay between study and test, the experiments of Chapter 4 tested whether there was a label-based bias in memory for hue when testing occurred a few minutes after exposure. A labeling effect was found but it was not in the direction predicted by RSH. Labeling the color category during the study phase of the experiment resulted in less biased memory than making a judgment unrelated to color. At least in the domain of hue, it appears that the RSH does not account for memory bias patterns.

For Chapter 3, I, along with my co-author Evan Heit, suggested alternative explanations for the patterns of results. The atypical bias observed in Chapter 3 could be a distinctiveness bias with immediate processing privileging the more unique aspects of a stimulus. With the immediate same/different judgments, response patterns reflect a higher standard for issuing a same judgment for hues that are more likely of the category and less distinct in the perceivers' experience. Additionally with the immediate testing, it may not have been beneficial to give much attention to the color categories when current experience made it clear that distinctions were being made at a level far below the basic color categories of red and green.

These explanations are not strongly representational or anti-representational. Little emphasis has been put on the contents of the mind other than some version of memory for visual experiences that are no long available in the environment. Most cognitive scientists would likely refer to this as representation yet this behavioral pattern does not require rigid object-like representation. Insofar as anti-representational theories have an account of working-memory-like processes, they could probably account for this simple memory usage. With regards to embodiment, essentially we have suggested that performance on the task is context dependent, or a product of the agent-environment system. The task demands of the fine hue distinctions and of the immediate memory test influence the observed memory bias more than the labeling manipulation. These task demands are relevant to the activity the participants were engaged in. Labeling the hues on top of the memory activity was superfluous to that activity. In fact, it could be that when labeling is more relevant to the task, it has a more prominent effect on memory bias. That, however, is a question left to future research. In accord with the RSH, these new hypotheses are about a task that has little bodily involvement. The influence of the agent-environment system and the history of participants in their past environments creating systematic behavioral biases are the main connection to embodiment for our hypotheses.

With no evidence of language creating bias at encoding in the experiments of Chapter 3, an alternative reason for a prototypical bias in response pattern is category knowledge affecting memory biases at memory retrieval. Therefore, in Chapter 4, I, again along with my co-author Evan Heit, suggest an explanation for our results using the logic of the Category Adjustment Model (CAM; Huttenlocher et al., 1991, 2000, 2007). Essentially, the initial memory trace is encoded with some degree of detail. The more strong an individual memory trace is, the less biased it will be. The weaker the memory trace, the more that recognition of an item will depend on other knowledge filling the knowledge gap. In Chapter 4, regardless of labeling, there was a bias toward selecting hues more typical of the color category at test. There was an effect of labeling such that labeling hues with their basic color category at encoding led to less biased recognition memory than alternative activities to labeling at encoding. The labeling activity could have increased the strength of the hue aspect of the memory trace at encoding. To test this hypothesis, participants were warned of the upcoming hue memory test, creating task demands emphasizing the importance of the color component of the objects being labeled by color or being judged as living/non-living. The difference disappeared with all memory still biased toward typicality, with the magnitude of bias comparable to the previous memory biases observed in the labeling condition without forewarning of the memory task. The CAM is strongly representational as written, discussing memory traces as object-like encodings. Our hypothesis explaining the present results, inspired by the CAM, is not necessarily as strongly representational. As in the explanation for the Chapter 3 results, some notion of memory is needed. Otherwise, the behavior of varying strength of that memory is agnostic to representationalism. An anti-representational account could focus on the relative strength of affordances suggested by the color labeling vs non-color labeling tasks at encoding. Labeling colors suggests that color may be relevant for future action. When told color would be relevant for future action, participants performed equivalently to the labeling condition in the surprise memory test version of the experiment. As asserted in Chapter 2, exemplar theories are compatible with the aspect of embodiment that is concerned with the history of a cognitive agent in its environment. The history of using labels suggest they are generally used when there is a purpose for them such as drawing attention to a specific aspect of current perception. The history of seeing reds and greens suggests some hues are more likely to be seen. The biases observed in the Chapter 4 experiments would be adaptive to future experience that aligns with this history. Therefore, while there is no explicit test of body-based influences in this task, the hypothesis we put forth is based on influences within the full agent-environment system and in that sense is an embodied hypothesis.

Across the two chapters, context, whether labeling or other task demands, had a measurable influence on memory biases. Our interpretations of the results are not strongly representational though are couched in inherited representational language. Our interpretations are compatible with all forms of embodiment, most strongly with stances that view the history of an agent as the major source of conceptual knowledge and behavioral tendencies.

7.2.2 Event Segmentation Decisions

The Event Segmentation Theory (EST) is an explicitly internal-only representational theory and is only compatible with the conceptualization hypothesis of Shapiro's three embodiment categories. This was discussed in some detail in Chapter 5, Section 3. Briefly, the EST treats working memory event models - active representations - as representational objects that are built quickly from current experience mixed with expectations from long-term memory. These representational objects are active but held apart from new perception. When no longer accurate to incoming perceptual input, the current event model is stored in memory and a new discrete model is built. Conceptualization embodiment comes from event models being built in part from the current perceptual input from the perceiver's point of view.

In Chapter 5, I presented a critique of the EST that while people may be able to describe events by breaking them down into component parts, the evidence for perception being discretized as an ongoing automatic process is not convincing. Performance on the active event segmentation task, indicating the end of self-defined activity units with a keypress, is only somewhat related to any passive and automatic cognitive process of segmentation that may be occurring within a perceiver. Specifically, the EST puts a strong emphasis on the increase of prediction error beyond some threshold as the instigator of segmentation. However, in most segmentation experiments, participants see a video twice to segment it at two timescales. On the second viewing, little of the video should produce perceptual prediction error - the participant has just watched the exact video minutes before and has knowledge of what to expect next. Yet, participants are able to sensibly segment the videos on both the first and second viewing. Something other than perceptual prediction error must be motivating segmentation responses.

The two experiments reported in Chapter 6 are investigating the nature of the event segmentation task, focusing on segmentation decisions. In the task, participants are being asked to make explicit judgments about the structure of the activity stream they are perceiving. Both experiments produced discrete segmentation response patterns that align with previous research. The experiments reported here utilized continuous response measures in addition to or instead of keypresses. The continuous response was collected via action dynamics from moving a slider marker within its track. In Experiment 1, the slider was used to report continuous expectation of the current activity unit ending, i.e., the trigger for giving a discrete keypress response. Expectation values at segmentation were highly variable. If the event segmentation behavior reflects perceptual prediction error, participants should not expect the end of end of activities and should predominantly report low expectation went they give a segmentation response. Instead, over a third of segmentation responses were paired with a high expectation value in the top 20% of the expectation range with the rest of the responses evenly distributed across the range. Since participants could expect their own responses, the task does not purely reflect a perceptual segmentation process based on prediction error. In Experiment 2, the segmentation response itself occurred via the slider. If the task reflects prediction error-based segmentation, responses should exclusively have ballistic movement pattern rather than gradual or stepwise movements that demonstrate anticipation of the end of the activity unit. A person cannot simultaneously anticipate the end of an activity and have a large prediction error while perceiving the anticipated activity ending. However, participants engaged in a diverse array of response strategies and response movement patterns. Over half of trials also demonstrated at least one sub-threshold response that was initiated but did not culminate in a complete movement to the response location on the slider; Not all perceived activity ends resulted in segmentation responses. The event segmentation task is a reasoning and decision-making task using knowledge of activities, including predictions of ends of activities to select response points.

The event segmentation task behavior is the product of event perception and reasoning. It does not provide evidence about perceptual mechanisms. Event segmentation behavior, instead, can provide insight into what meaningful activities are, and how people explicitly conceive of activities. Abandoning the EST interpretation of the task, systematic behavior appears to be representation-hungry—outside knowledge of activities guides reasoning about activity units and unit end points. The design of these experiments using a continuous measure to capturing action dynamics requires a theoretical commitment to movement through space and time reflecting cognitive processes; an embodiment stance. Indeed, anticipation strategies exhibited by participants in Experiment 2 demonstrate an active process of utilizing the physical environment adaptively. The experimental results do not distinguish amongst theoretical stances on representation and embodiment; yet by being potentially consistent with a wide range of theories, the experiments demonstrate that the EST assumption of strong representation, and therefore limited embodiment, are not well justified.

7.2.3 Across Phenomena

Research in the cognitive sciences has long been interpreted through a strongly representational lens by default with the representations being amodal and disembodied. Representational shifts and event segmentation are proposed cognitive processes that are borne of representational theories. The major theories associated with these phenomena are not of that traditional sort. The RSH and the EST each branch away from the computer metaphor albeit in different ways. The RSH, while representational, is dynamically representational, viewing representations as malleable, at least during encoding. Different varieties of information in the environment have differing effects on cognitive processing, reflecting an important role of the agent-environment system in cognition. The EST is more traditionally representational with snap-shot unchangeable event models, essentially representational objects. These representations are influenced by a perceiver's body and history of experience. The event models contain scenes and modality-specific information about the world. The event model creation is driven by predictions based on previous experience of the environment from a perceiver's own viewpoint. These steps away from the traditional computer metaphor view of representation and disembodiment are interesting and exciting.

In this dissertation, I have questioned the aspects of these theories that adhere to traditional theoretical stances. Even theorizing about phenomena that appear to be representation-hungry, e.g., memory for specific items and identifying recurring structure in continuous sensory experience, do not require a traditional notion of representation. As

Connell and Lynott (2014) asserted, “you can't represent the same concept twice.” If representing is activating some collection and sequence of neural patterns, representing a concept for second time would require having the exact same activation as the first time. This is simply not possible given how complex the brain is. Some aspects of the previous activation pattern will not reactivate while other patterns will happen to be active, all it takes is a small difference in the environment, or a difference in what is primed by previous thoughts.

Representation is a useful construct when talking about brain-internal processes. Memory for things like objects and color hues as well as autobiographical experience of events is removed in space and time from the perceptual experience. Brain-based processes need to implement this memory storage. But calling a memory trace a representation and treating it as a cohesive entity misses the complexity and continued malleability of memory. Dynamic ideas of representation acknowledge that there is some cohesiveness to what is called a memory but also acknowledge that a myriad of factors introduce variability into cognitive processing, including memory processing. The empirical results of this dissertation call for dynamic views of memory storage. Memory for hue half a second after it is experienced is not veridical to experience (Chapter 3). Accuracy of memory for hue after minutes depends on the context in which the original hue was experienced (Chapter 4). Event segmenting behavior is highly variable in quantity and manner for different participants watching the same sequence of activity (Chapter 6). These are examples of how traditional representation is not adequate for explaining memory. Theorizing clean, contained representations between imperfect sensory input and variable action output would require the brain to be doing 100% of the work for a fraction of the payoff. Given the high variability observed in behavior related to memory, representations (or relatively consistent processing patterns, if you prefer) are more appropriately conceived of as fuzzy, impressionistic, and variable with context.

Both phenomena of interest have been viewed as embodied in a minimal sense, with the history of the cognitive agent in its environment essential to the theories in the literature as well as the hypotheses I have put forward. The context of sitting in a behavioral laboratory’s running room and being given strange instructions on a computer screen is an embodied experience. Being asked to label hues has pragmatic meaning that has consequences for subsequent memory of the hues. Asking participants to mark the smallest or largest “natural and meaningful activity units” while watching a video is not an isolated instruction; the instructions “smallest” and “largest” produce a context of comparison, introducing task demands. Embodiment may or may not be a paradigm shift. Regardless of embodiment’s paradigmatic status, awareness that all human experience occurs within a body and within a history is not just of theoretical importance but also practical importance to researchers regardless of their individual theoretical leanings.

7.3 Broader Implications

7.3.1 Object Metaphors

I opened this dissertation with a discussion about the use of object metaphors in cognitive science and how they can imply the associations beyond what the communicator intends. This dissertation itself is riddled with object metaphors. There are memory *traces*; *percepts*; event *segments, models,* and *schemata*; and category *exemplars* and *prototypes.* Representations *shift*; processing has *depth*; memory has *strength,* is *stored* and *retrieved*; and perception can be *distorted* or *lost.* This is the language of the field and in particular it is almost impossible to talk about memory without relying on object metaphors - how can a process stop and start without being stored in between? But perhaps the object permanence implied by these metaphors is why it seems so strange to think of memory as flexible and, in any particular instantiation, ephemeral. These metaphors undoubted influence the way cognitive theories and experiments designed.

The phenomena under investigation in this dissertation cannot be discussed without some object metaphors for cognitive processes. Representational shifts are a phenomena named with an object metaphor. When possible, I discussed response bias and shifts in responses within color space rather than shifts in memory itself. However, the RSH is about shifts along a typicality gradient within a category space; a metaphorical mental space with psychologically defined dimensions. The EST's proposed event segmentation perceptual process produces event models which are built and stored. Discussing the EST requires use of these metaphors.

Should embodiment theories and nuanced views of representation continue to gain more support, the language of cognitive science may become less object oriented. By discussing the issues and results through the lens of these theoretical domains, I hope I have conveyed that cognition is better thought of as processes and the implications of object metaphors such as discrete rather than fuzzy bounds are assumptions to be questioned wherever they arise.

7.3.2 Perception and Memory in Context

From the theoretical discussions of agent-environment systems, mutuality relations, and affordances; to the effect of pragmatic task demands in the representational shift experiments; and to the influence of existing knowledge on event segmentation responses, one of the key takeaways of this dissertation is that context matters.

As discussed in the introduction, perception is affected by existing knowledge, and memory is often affected by newly perceived information. In psychophysics, it has long been known that concurrent experience can change how sensory signals are perceived, e.g., the McGurk effect (McGurk & MacDonald,1976) and the checkershadow illusion (Adelson, 1995). In the representational shift experiments, concurrent perception of labels while perceiving hues did not result in the hybrid memory biased toward the labeled category as suggested by the RSH. Pragmatically, the labeling did not have a clear purpose to the participant. However, usually, language does have a functional purpose. When a labeling effect was observed in the Chapter 4 experiments, the labeling of color seemed to be a cue that the color was more important than other aspects of the image. So rather than perceptual context affecting the perception-to-memory processing stream, it was instead the extra-perceptual pragmatic context that appears to have driven the effect. Similarly, in the Chapter 3 experiments, though there was not an effect of labeling, the sequence of rapid exposure then test may have driven the perception-to-memory stream to focus on distinctive information.

Perceptual prediction error cannot be the only source of information for event segmentation decisions. When videos are watched a second time by the same participant, prediction error for the incoming perceptual signal should be considerably lower than it was during the initial exposure. Given instructions to mark off natural and meaningful activity units, participants seem to be making segmentation decisions based on their understanding of the overarching activity rather than perceptual unpredictability. Rather than reflecting a discrete, online working memory event modeling process, the task demonstrates how perception and memory align and mutually influence behavior in a particular behavioral context.

Task demands such as those observed to be so influential in the present experiments are often viewed as noise in experimental data analysis. The full context of an experiment from the physical environment, to the pragmatic implications of experimental design choices, and to experiment external factors such as conducting the experiment at a particular time of year can all influence participant behavior. Taking into account the full context and designing experiments that capture rather than obscure external influences on perception and memory are essential for gaining an accurate view of the processes and their interactions.

7.3.3 Event Cognition

To make a bold claim, in many ways event cognition is cognition. Humans and other organisms perceive their environment extended over time. Even experiencing something static is an experience over time, one lacking variability. More specifically for the behavioral sciences, experiments are events that participants take part in. Therefore all behavioral scientists should have some concern for how event cognition works if only to better understand how participants interact with the tasks they design. In spite of this, event cognition is relatively young as an area of research.

I’m not convinced that segmentation is a natural part of perception. If some version of segmentation is a core aspect of perception, it is not likely to be as discrete as the EST & the related Event Horizon Model (EHM; Radvansky & Zacks, 2014) make it out to be. Therefore, the name segmentation is likely to be a mis-applied metaphor. Instead, event understanding could be the result of continuous information processing with constant sensory input and spreading neural activation through association networks. Periodically, we transition from one understanding to another. In complex systems, small signals can accumulate moving the state of the system and if a threshold is passed the movement can become much more rapid until another stable state is reached. Such a transition may look discrete from a coarse-grained view of the system but at a fine-grain, it is nonlinear continuity. That continuity has consequences that look like noise in the system if treated as sharp discrete processes. More than most areas of cognitive science, event cognition is grounded in the interaction between the perceiver and its environment. Some ecological psychologists (Chemero, 2000; Chemero, Klein, & Cordeiro, 2003) go so far as to suggest that events explicitly are changes in affordances, opportunities for action with in the environment. Regardless of whether that stance is accepted broadly, taking the agent-environment system seriously is necessary for interpreting event perception and other related cognitive processes. Event cognition is an area of cognitive research that requires embodied theories and, in my opinion, the conceptualization version of embodiment is not enough.

7.4 Giving Responses Dimension

A key methodological theme of this dissertation, as the title *Giving Responses Dimension: Representational Shifts in Color Space and Event Segmentation Decisions in Physical Space Over Time* asserts, is introducing changes to existing experimental paradigms to garter more information than previous versions had collected using increase dimensionality.

With the representational shift experiments of Chapters 3 and 4, rather than using target stimuli and matched foils that are arbitrarily similar, the experiments presented here used targets and foils that were similar in a systematic fashion along a single measurable dimension. The task remained a comparison task requiring discrete responses, but the pattern of responses created sensitivity metrics that spanned hue space, and collapsed across targets, spanned typicality space with some foils more typical of the category and some foils less typical of the category. Beyond the dimension of color space and its category structure, the combined experiments of Chapters 3 and 4 produce a view of representational shifts over the dimension of time with varying study to test delays of half a second, five seconds, and a few minutes. The delay differences give us a view of processing extended over time.

With event segmentation, previous research had collected discrete segmentation responses at time points along the activity sequence. The Chapter 6 experiments also collected these same segmentation points. The continuous slider responses, introduced for the first time to this paradigm in this dissertation, give segmentation responses additional dimension. In Experiment 1, the slider response provides an ongoing recording of the expectation of the end of the current event, providing a continuous dimension along side the binary segmentation time series. In Experiment

2, the slider response extends the previously binary response over time and space. The responses could be slow or fast, and smooth or stepped, including pauses. Aborted responses are a signal that is not observable with the binary response but with the slider can be seen in the form of moving toward the response end without reaching the segmentation response location.

In my presented empirical investigations, using dependent measures that capture more detail have provided new insight into the phenomena of interest. Expanding discrete responses into responses with information along a measurable dimension is an important strategy in research that looks to question assumptions.

7.5 Conclusion

Every individual is part of a unique cognitive system; a homegrown complex system of capabilities and past experiences. Organisms are the product of a unique trajectory of experience in the world. Each human is born with a certain baseline of biological components with some variability from other humans. After that, the experience of each human is different. Even identical twins who have identical DNA will make and learn from different mistakes, have different interpersonal interactions, and develop interests that diverge. Even when in the same overall context, their experiences will differ slightly. Perhaps the twin who has a first name that is earlier in the alphabet will be regularly asked to speak before the other in school. One twin may take on a leader role within their relationship creating a systemic difference with how they interact with each other and with others while together. No matter how similar two people's lives are, their interaction with the environment will differ resulting in unique neural networks through which incoming sensory information is processed.

Understanding the manner and mechanisms by which humans and other cognitive agents make sense of the world has been a long-standing quest in the cognitive sciences. Reacting to and interacting with the environment a person finds him or herself in is a fundamental cognitive capacity. Meaning within the brain emerges from a complex system of neuronal pathways with degrees of freedom beyond imagining; from neuronal firing rates to their synaptic connections to the influence of electric oscillations produced by neural firing to the concentrations of neurotransmitters and other molecules, the signals and signal interactions are staggering. With the inclusion of the environment and all the affordances both local, e.g., picking up a glass or flipping a light switch, and extended, e.g., calling someone to act in your stead or even getting on a flight that eventually results in a flipped a light switch thousands of miles away from the initial location, the system producing meaning is amazingly complex.

Cognitive scientists cannot account for all of the moving parts of a cognitive system simultaneously. By being aware of the complexity and remaining vigilant about questioning assumptions, scientists can identify unexpected influences in their investigations. Throughout this dissertation

I have sought to question assumptions at many levels. In the introduction, I began by acknowledging that the language of science can bias scientists, myself included. I explicitly investigated and questioned theoretical assumptions in the domains of representation and embodiment for the two phenomena under empirical study. Finally, I questioned whether the empirical evidence for existing theories truly supports their claims. My empirical designs challenge the assumption that binary responses are informative enough, pushing deeper through adding dimension to previously discrete tasks. Theoretically driven investigations such as those I have reported here advance cognitive science by challenging existing theories, supporting them or helping to recontextualize existing research.

References

Abbot, V., Black, J. H., & Smith, E. E. (1985). The representation of scripts in memory. *Journal of Memory and Language*, *24*, 179-199.
Adelson, Edward H. (2005). "Checkershadow Illusion".
Abrams, R. A., & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science*, *2*(3), 153-157.
Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261-295.
Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(5), 1063-1087. doi: 10.1037/0278-7393.20.5.1063
Baars, B. J. (1986). *The Cognitive Revolution in Psychology.* New York: Guilford Press.
Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General.* Advance online publication.
Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709-721.
Balota, D. A., & Abrams, R. A. (1995). Mental chronometry: beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1289.
Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The “reading the mind in the eyes” test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology & Psychiatry*, *42*, 241-251.
Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition*, *4*, 311-331.
Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255-278.
Barrett, L. (2011). *Beyond the Brain: How Body and Environment Shape Animal and Human Minds.* Princeton, NJ:Princeton University Press.
Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577-609.
Barsalou, L. W., Santos, A., Simmons, W. K., and Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. de Vega, A. Glenberg, & A. Graesser (Eds.) *Symbols and Embodiment: Debates on meaning and cognition* (245-284). Oxford University Press: Oxford.
Barsalou, L. Simmons, W. K., Barbey, A., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84-91.

Bartlett, J. (2014, August 17). Robustness of linear mixed models [Blog post]. Retrieved from http://thestatsgeek.com/2014/08/17/robustness-of-linear-mixed-models/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Blanco, N. and Gureckis, T. M. (2013). Does category labeling lead to forgetting? *Cognitive Processing, 14,* 73-79.

Berlin, B., & Kay, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley: University of California Press.

Boltz, M. (1992). Temporal accent structure and the remembering of filmed narratives. *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 90-105.

Bower, G. H., & Rinck, M. (2001). Selecting one among many referents in spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 81-98.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*, 981-990.

Brooks, R. (1991a). New approaches to Robotics. *Science*, *253*, 1227-1232.

Brooks, R. (1991b). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellait, B., & Williamson, M. (1999). The cog project: Building a humanoid robot. In C. Nehaniv (Ed.), *Computation for Metaphors, Analogy, and Agents. Lecture Notes in Artificial Intelligence 1562* (52-87). New York: Springer.

Bukach, C. M., Gauthier, I., & Tarr, M. J. (2006). Beyond faces and modularity: The power of an expertise framework. *Trends in Cognitive Sciences*, *10*(4), 159-166. doi: 10.1016/j.tics.2006.02.004

Cesario, J. (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, *9*(1), 40-48. doi: 10.1177/1745691613513470

Chemero, A. (2009). *Radical Embodied Cognitive Science.* Cambridge, MA: MIT Press.

Chemero, A., Klein, C., & Cordeiro, W. (2003). Events as changes in the layout of affordances. *Ecological Psychology*, *15*(1), 19-28.

Chemero, A. (2000). What events are. *Ecological Psychology*, *12*(1), 37-42.

Chomsky, N. (1975) *Reflections on Language.* New York: Pantheon Books.

Chomsky, N. (1957). *Syntactic Structures.* The Hague: Moutin.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy, 78*(2), 67-90. doi: 10.2307/2025900

Churchland, P. M. (1988). Perceptual plasticity and theoretical neutrality: A reply to Jerry Fodor. *Philosophy of Science, 55*, 167–187.

City University (2002). A web-based colour vision test. http://www.city.ac.uk/avrc/colourtest.html.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7-19.

Clark, A., & Toribio, J. (1994). Doing without representing? *Synthese*, *101*, 401-431.

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: an R package. *Frontiers in psychology*, *5*, 510.

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*, 390-406.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. doi: 10.1037/0096-3445.104.3.268

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15-28.

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioral Brain Research, 76,* 181-189. doi: 10.1016/0166-4328(95)00198-0

Firestone, C., & Scholl, B. J. (2015). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *20*, 1-77.

Fischhoff, B. (1975). Hindsight ≠ foresight: the effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299.

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Gauthier, I., & Logothetis, N. K. (2000). Is face recognition not so unique after all?. *Cognitive Neuropsychology*, *17*(1-3), 125-142. doi: 10.1080/026432900380535

Gauthier, I., & Nelson, C. A. (2001). The development of face expertise. *Current Opinion in Neurobiology*, *11*, 219-224.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191-197. doi:10.1038/72140

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, *17*, 324-363.

Gibbs, R. W. (2006). *Embodiment and Cognitive Science.* Cambridge, UK: Cambridge University Press.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Boston: Houghton-Mifflin.

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin and Review*, *9*, 558-565.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*, 69-83.
Glenburg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, *8*, 573-585.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251-279.
Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychological Bulletin and Review*, *23*, 959-978.
Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*, 178-200.
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology*, *49*(1), 585-612. doi: 10.1146/annurev.psych.49.1.585
Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69-78. doi: 10.1002/wcs.26
Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition as soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152-160. doi: 10.1111/j.0956-7976.2005.00796.x
Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.) *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 843). Austin, TX: Cognitive Science Society.
Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin & Review, 18*, 355-363.
Hard, B. M., Tversky, B., & Lang, D. S. (2006). Making sense of abstract events: Building event schemas. *Memory & Cognition*, *34*, 1221-1235.
Harnad, S. (1987). *Categorical Perception: The Groundwork of Cognition.* Cambridge: Cambridge University Press.
Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*, 301-307.
Hayes, B. K., Heit, E., & Rotello, C. M. (2014). Memory, reasoning and categorization: Parallels and common mechanisms. *Frontiers in Psychology* *5*, 529. doi: 10.3389/fpsyg.2014.00529
Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology, 24*, 341-380.
Heit, E., (1994). Models of the effects of prior knowledge on category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 1264-1282.
Heit, E. (1997). Knowledge and concept learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts, and Categories*, 7-41. Psychology Press.

Heit, E., & Barsalou, L. (1996). The instantiation principle in natural categories. *Memory, 4*, 413-451.
Heit, E. & Hayes, B. K. (2005). Relations among categorization, induction, recognition, and similarity. *Journal of Experimental Psychology: General, 134*, 596-605. doi: 10.1037/0096-3445.134.4.596
Heit, E., & Rotello, C. M. (2014). Traditional difference-score analyses of reasoning are flawed. *Cognition, 131*, 75-91.
Heit, E., Rotello, C. M., & Hayes, B. K. (2012). Relations between memory and reasoning. In Ross, B. H. (Ed.), *Psychology of Learning and Motivation, 57*, 57-101. Academic Press.
Hendrickson, A. T., Carvalho, P. F., & Goldstone, R. L. (2012). Going to extremes: The influence of unsupervised categories on the mental caricaturization of faces and asymmetries in perceptual discrimination. *Proceedings of the Thirty-Fourth Annual Conference of the Cognitive Science Society,* (pp. 1662-1667). Sapporo, Japan: Cognitive Science Society.
Hill, D. E. (1979). Orientation by jumping spiders of the genus Phidippus pulcherrimus (Araneae: Salticidae) during the pursuit of prey. *Behavioral Ecology and Sociobiology*, *5*, 301-322.
Hintzman, D. L. (1984). Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers, 16*, 96-101.
Hsieh P. J., Vul E., & Kanwisher N. (2010). Recognition alters the spatial pattern of FMRI activation in early retinotopic cortex. *Journal of Neurophysiology. 103*(3), 1501-1507. doi: 10.1152/jn.00812.2009
Huttenlocher, J. Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototypical effects in estimating spatial location. *Psychological Review, 98*, 352-376.
Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment?. *Journal of Experimental Psychology: General*, *129*, 220.
Huttenlocher, J., Hedges, L. V., Lourenco, S. F., Crawford, L. E., & Corrigan, B. (2007). Estimating stimuli from contrasting categories: Truncation due to boundaries. *Journal of Experimental Psychology: General, 136,* 502-519.
Jackson, R. R. & Wilcox, R. S. (1990). Aggressive mimicry, prey-specific predatory behaviour and predator recognition in the predator-prey interactions of Portia fimbriata and Euryattus sp. jumping spiders from Queensland. *Behavioral Ecology and Sociobiology*, *26*, 111-119.
Jackson, R. R. & Wilcox, R. S. (1994). Spider flexibly chooses aggressive mimicry signals for different prey by trial and error. *Behavior*, *127*, 21-36.
Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326-338.
Jeffreys, H. (1961). *Theory of probability*. Oxford: UK Oxford University Press.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialised for face perception. *Journal of Neuroscience, 17*, 4302–4311.
Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis?. *American Anthropologist, 86,* 65-79.
Kelly, L. J. & Heit, E. (2013). Representational shifts towards the prototype in memory for hue. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceeding of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
Kelly, L. J. & Heit, E. (2014). Representational shifts made visible: Movement away from the prototype in memory for hue. *Frontiers in Psychology, 5*.
Kelly, L. J., & Heit, E. (2017). Recognition memory for hue: Prototypical bias and the role of labeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 955.
Khemlani, S. S., Harrison, A. M., & Trafton, J. G. (2015). Episodes, events, and models. *Frontiers in human neuroscience*, *9*, 590.
Kikutani, M., Roberson, D., & Hanley, J. R. (2010). Categorical perception for unfamiliar faces: The effect of covert and overt face learning. *Psychological Science, 21*(6), 865-872.
Konkle, T., Brady, T. F., Alvarez, G. A., & Olivia, A. (2010). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21,* 1551–1556. http://dx.doi .org/10.1177/0956797610385359
Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.
Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Science*, *12*, 72-79.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26.
Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, *33*, 1-64.
Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By.* Chicago: University of Chicago Press.
Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought.* New York: Basic Books.
Lawrence, Z., & Peterson, D. (2016). Mentally walking through doorways causes forgetting: The location updating effect and imagination. *Memory*, *24*, 12-20.
Liberman A. M., Harris K. S., Hoffman H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368. doi: 10.1037/h0044417
Levin, D. T., Beale, J. M. (2000). Categorical perception occurs in newly learned faces, other-race faces, and inverted faces. *Perception & Psychophysics, 62*, 386–401. doi: 10.3758/BF03205558

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of verbal learning and verbal behavior*, *13*(5), 585-589.
Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin & Review*, *15*, 838-844.
Louwerse, M. M. (2010). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273-302.
Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*, 96-104.
Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R. D. & Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software].
Lupyan, G. (2008). From chair to "chair:" A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General, 137*, 348-369. doi: 10.1037/0096-3445.137.2.348
Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*, 54. doi: 10.3389/fpsyg.2012.00054
Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Labels facilitate learning of novel categories. *Psychological Science, 18*, 1077-1083. doi: 10.1111/j.1467-9280.2007.02028.x
Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implication. *Current Biology, 23*, 2023-2027.
Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Second edition. Psychology Press.
Mahon, B. Z. (2014). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*, 420-429.
Mahon, B. Z. (2015). Response to Glenberg: Conceptual content does not constrain the representational format of concepts. *Canadian Journal of Experimental Psychology*, *69*, 179-180.
Mahon, B. Z. (2015). The burden of embodied cognition. *Canadian Journal of Experimental Psychology*, *69*, 172-178.
Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, *102*, 59-70.
Mandler, J. M. & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic?. *Cognitive Development, 3,* 247-264.
McGurk H., MacDonald J. (1976). "Hearing lips and seeing voices". *Nature*. **264** (5588): 746–8.
McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces?. *Trends in Cognitive Sciences*, *11*(1), 8-15. doi:10.1016/j.tics.2006.11.002
Meyer, M., & Baldwin, D. (2011). Statistical learning of action: The role of conditional probability. *Learning & Behavior*, *39*, 383-398.

Mitterer, H., Horschig, J. M., Musseler, J., & Majid, A. (2009). The influence of memory on perception: It's not what things look like, it's what you call them. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1557–1562. http://dx.doi.org/10.1037/ a0017019

Morris, C. D., Bransford, J. D. & Franks, J. J. (1977) Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533. doi:10.1016/S0022-5371(77)80016-9

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. *Journal of Memory and Language*, *28*, 292-312.

Morrow, D. G., Greenspan, S. L., & Bower, G. H. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language*, *26*, 165-187.

Nadel, L., Hupbach, A., Gomez, R., & Newman-Smith, K. (2012). Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1640-1645. doi: 10.1016/j.neubiorev.2012.03.001

Nader, K., & Einarsson, E. Ö., (2010). Memory reconsolidation: an update. *Annals of the New York Academy of Sciences*, 27-41.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*, 113-126.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, *28*(1), 28.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115,* 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700-708.

Nosofsky, R. M. (2015). An exemplar-model account of feature inference from uncertain categorizations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 1929-1941.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*, 939-973.

Palmeri, T. J., & Nosofsky, R. M., (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology, 54,* 197-235.

Pérez-Carpinell, J., Baldoví, R., de Fez, M. D., & Castro, J. (1998). Color memory matching: Time effect and other factors. *Color Research and Application, 23,* 234–247.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition, 118*, 306-338.

Persaud, K., & Hemmer, P. (2014). The influence of knowledge and expectations for color on episodic memory. *Proceedings of the Thirty-Sixth Annual Conference of the Cognitive Science Society*, 1162-1167. Quebec City, Canada: Cognitive Science Society.

Pettijohn, K. A., & Radvansky, G. A. (2016). Narrative event boundaries, reading times, and expectation. *Memory & Cognition*, *44*, 1064-1075.
Pettijohn, K. A., Thompson, A. N., Tamplin, A. K., Krawietz, S. A., & Radvansky, G. A. (2016). Event boundaries and memory improvement. *Cognition*, *148*, 136-144.
Pinker, S. (1994). The Language Instinct: How the Mind Creates Language. New York, NY: Harper Perennial Modern Classics.
Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead. Long live the sleeper effect. *Journal of Personality and Social Psychology, 54*, 203-218.
Proffit, D. R. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science*, *1*, 110-122.
Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*, 341-365.
Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*, 1102-1107.
Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, *34*, 1150-1156.
Radvansky, G. A., Krawietz, S. A., & Tamplin, A. K. (2011). Walking through doorways causes forgetting: Further explorations. *The Quarterly Journal of Experimental Psychology*, *64*, 1632-1645.
Radvansky, G. A., Tamplin, A. K., & Krawietz, S. A. (2010). Walking through doorways causes forgetting: Environmental integration. *Psychonomic Bulletin & Review*, *17*, 900-904.
Radvansky, G. A., & Zacks, J. M. (2014). *Event Cognition.* Oxford: Oxford University Press.
Regier, T. & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Science, 13,* 439-446. doi: 10.1016/j.tics.2009.07.001
Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*, 1436-1441.
Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75.
Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*, 613-643.
Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive psychology*, *19*(4), 473-497. doi: 10.1016/0010-0285(87)90016-8

Richler, J.J., Gauthier, I., & Palmeri, T.J. (2011). Automaticity of basic-level categorization accounts for labeling effects in visual recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1579-1587.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2013) How does using object names influence visual recognition memory? *Journal of Memory and Language, 68*, 10-25.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition*, *28*, 1310-1320.

Rinck, M., & Weber, U. (2003). Who when where: an experimental test of the event-indexing model. *Memory & Cognition*, *31*, 1284-1292.

Roberson, D., & Hanley, J. R. (2010). Relatively speaking; An account of the relationship between language and thought in the color domain. In B. Malt & P. Wolff (Eds.), *Words and the World: How words capture the human experience,* (pp. 183-198). Oxford University Press.

Roberson, D., Hanley, J. R., & Pak, H. (2009). Thresholds for color discrimination in English and Korean speakers. *Cognition*, *112*, 482-487.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Rosch, E. (1975). The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(4), 303-322. doi: 10.1037/0096-1523.1.4.303

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*, 382-439.

Ross, D. (1951). *Plato's Theory of Ideas.* Oxford: Clarendon Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926-1928.

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology*, *18*, 37-55.

Schwanenflugel, P. J., & Akin, C. (1994). Developmental trends in lexical decision for abstract and concrete words. *Reading Research Quarterly*, *29*, 251-263.

Searle, J. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, *3*, 417–457

Shaprio, L. (2011). *Embodied Cognition.* New York: Routledge.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156–163. http://dx.doi.org/10.1016/s0022-5371(67)80067-7

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Science, 4*(4), 147-155. doi:10.1016/S1364-6613(00)01455-8

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*, 246-251.

Sloutsky, V. M., & Fisher, A. V. (2012). Linguistic labels: Conceptual markers or object features?. *Journal of Experimental Child Psychology*, *111*(1), 65-86. doi: 10.1016/j.jecp.2011.07.007
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34-49.
Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, *20*, 989-999.
Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, *53*, 125-140.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*, 1-29.
Spivey, M. (2008). *The continuity of mind*. Oxford University Press.
Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(29), 10393-10398.
Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, *253*, 1380-1386.
Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149. doi: 10.3758/BF03207704
Sturges, J. & Whitfield, T.W.A. (1995). Locating basic colours in the munsell space. *Color Research and Application, 20*, 364-376.
Swallow, K. M., & Zacks, J. M. (2004). Hierarchical grouping of events revealed by eye movements. *Poster presented at 45th annual Meeting of the Psychonomic Society, Minneapolis, MN.*
Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*, 236-257.
Tarsitano, M. (2006). Route selection by a jumping spider (Protia labiata) during the locomotory phase of a detour. *Animal Behaviour*, *72*, 1437-1442.
Tarsitano & Andrew (1999). Scanning and route selection in the jumping spider, Portia labiata. *Animal Behaviour*, *58*, 255-265.
Tomasello, M. (2009). *Constructing a Language.* Cambridge, Massachusetts: Harvard University Press.
Toveé, M. J. (1998). Is face processing special? *Neuron, 21*, 1239–1242. doi: 10.1016/S0896-6273(00)80644-3
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
Whorf, B.L. (1956). *Language, Thought and Reality*. J.B. Carroll (Ed.). Cambridge, Massachusetts: MIT Press.

Wilcox, R. S., Jackson, R.R., and Gentile, K. (1996). Spiderweb smokescreen: spider trickster uses background noise to mask stalking movements. *Animal Behaviour*, *51*, 313-326.

Wilson, A. D., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology*, *4*, 58.

Winawer, J., Witthoft, N., Frank, M.C., Wu, L., Wade, A.R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy Sciences, 104*, 7780 –7785.

Zacks, J. M. (2004). Using movement and intentions to understand simple events. *Cognitive Science*, *28*, 979-1008.

Zacks, J. M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., Buckner, R.L., & Raichle, M.E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651-655.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin*, *133*, 273-293.

Zacks, J. M., Speer, N. K., Vettel, J. M., & Jacoby, L. J. (2006). Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging*, *21*, 466-482.

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*, 150-162.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*, 29-58.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation*, *44*, 35–62. New York: Academic Press.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 386-397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological bulletin*, *123*(2), 162.

Zwaan, R. A., Radvansky, G. A., Hillard, A. E. & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies in Reading, 2*, 199-220.