A Dual Model of Relational Concept Representation

by

Daniel Corral

B.A., California State University, Northridge, 2009

M.A., California State University, Northridge, 2011

M.A., University of Colorado Boulder, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2017

This thesis entitled:

A Dual Model of Relational Concept Representation

written by Daniel Corral

has been approved for the Department of Psychology and Neuroscience

Signature _____

Matt Jones (chair)

Signature _____

Alice Healy

Signature _____

Tim Curran

Signature _____

Michael Mozer

Signature _____

Michael Saddoris

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the

content and the form meet acceptable presentation standards of scholarly work in the above

mentioned discipline.

IRB protocol # _____16-0562_____

Corral, Daniel (Triple Ph.D., Cognitive Psychology, Cognitive Science, & Cognitive Neuroscience)

A Dual Model of Relational Concept Representation

Thesis directed by Associate Professor Matt Jones, Department of Psychology and Neuroscience

Relational concepts pervade daily life, as people are regularly required to comprehend, articulate, and reason about relational ideas and scenarios. Critically, these processes might be altered by how such concepts are represented. The dominant theories of relational learning have been built on the assumption that relational concepts are represented compositionally, based on the relationships among a concept's components. However, these theories have typically neglected the possibility that a concept's components can be consolidated or chunked into a unitized concept, producing a representation that is devoid of the concept's component parts. The distinction between compositional and unitary representations of relational concepts is a natural consequence of structure-mapping theory, but its psychological implications have not been explored. This paper reports 7 studies that examine how people represent relational concepts and how such representations affect relational learning. The general take away from these studies is that people do indeed appear to be capable of representing relational concepts in two fundamentally different ways, unitarily and compositionally. Furthermore, unitary representations seem to lead to better relational learning than compositional representations, especially for inference-based tasks. However, the data suggest that there might be various factors that interact with how representation affects relational learning (e.g., individual differences in representation, type of task, type of comparison). The conclusion that follows from these studies is that unitary representations might incur less cognitive load than structural

alignment of compositional representations, and thus may be the default for everyday relational reasoning.

Acknowledgements

- I would first like to start off by thanking my parents for providing me with the opportunities that have allowed me to get to this point, thank you both! I would also like to thank my three siblings for putting up with my incessant, silly philosophical arguments and ideas that I pushed over the years (sorry about that, but it helped to develop my critical thinking at an early age so thank you). Thank you to all of my friends who were willing to listen to my ramblings and engage me in debate over the years. I would also like to thank all of the friends I have made in graduate school for all of their support. Thank you all!

- I would like to thank Dr. Michael Saddoris and Dr. Mike Mozer for agreeing to be on my committee and reading through this somewhat long dissertation (sorry in advance). I also want to thank Dr. Tim Curran for being on my committee(s), as well as the detailed feedback and advice Tim has given me over the years. You have always been very supportive Tim and I really appreciate it, and thank you for taking the time to write me a letter of recommendation for my postdoc.

- I would like to give a big thank you to Dr. Jill Quilici and Dr. Abe Rutchick for their early mentorship and strong support. Jill was always very supportive and caring. The same can be said of Abe. Abe and Jill, you both played a vital role in my training and are in large part responsible for me getting into graduate school, without which I would not by writing this acknowledgment section now. Thank you both!

- I would like to give a great thank you to Dr. Alice Healy. Alice, you have always been incredibly supportive and have been a beacon of positivity. I greatly appreciate all of your feedback and enthusiasm. Your positive energy is contagious and you are great to work with. Thank you for being on all of my committees, for writing me letters of recommendations for my postdoc applications, and for the training I have received in your lab. I really appreciate everything you have done, thank you!

- Lastly, I would like to give the biggest thank you to my advisor, Dr. Matt Jones. Matt, you are a great advisor and I have been saying this for years now to anyone who would listen, but I have not told you: I am convinced that you are one of the best advisors in the world and I truly feel fortunate to have had the honor of working with you in graduate school. I greatly appreciate the feedback and training I have received from you. Thank you for the amazing patience you showed in my early years in graduate school and thank you for all of the time you spent helping me program over my first year and half or so. With the time you have dedicated to my training, you have gone well above and beyond what can be reasonably expected of any advisor. Thank you for always being supportive and for helping me to become the scientist that I am today. You have been the most influential figure in my training and whatever academic or scientific success I may achieve I will in great part owe to you. Thank you, Matt!!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

How do people represent conceptual information? For many years this question has vexed cognitive scientists and philosophers alike. It has been posited that concept representation is critical for many higher-order tasks, such as analogical reasoning, problem solving, decision-making, and category and concept acquisition (Markman, 1999; Markman & Dietrich, 2000). However, the specific role that representation plays in such tasks is often vague and underspecified. One reason for this lack of specificity is that peoples' representations cannot be readily accessed or measured. Compounding the problem is that a given concept can typically be represented in many different ways, making it challenging for researchers to gain definitive knowledge on the matter. Consider the simple example of a dog: A dog can be represented based on its component parts (e.g., tail, ears, fur, etc.,) or as an atomic attribute that represents the animal as a whole (i.e., the concept of *dog*). Recent work has referred to this latter type of representation as a *unitary concept*, in which the component parts of the concept are not explicitly accessed or represented (Corral & Jones, in preparation; Corral, Kurtz, & Jones, 2017).

It is important to note that there are numerous domains within cognitive science that formalize representation in various ways. The present paper operates within the framework of structure-mapping theory (Gentner, 1983), which makes the implicit assumption that a relational concept can be represented in two fundamentally different ways: (1) as a system of relations, with meaning derived both from the identities of those relations and from how they are interconnected by shared role-fillers (Corral & Jones, 2014); or (2) as a primitive, atomic relation that is explicitly represented. These two types of representations are referred to as *compositional*

and *unitary* representations, respectively. Although this logical distinction has been noted (Gentner, 1983), its potential psychological implications have largely been neglected.

To elaborate further, the first of these representational assumptions is premised on the idea that representations are constructed from two basic types of building blocks: objects and relations. The second assumption is based on the idea that a relation operates on a set of *n* objects, that is, for every ordered set of *n* objects, the relation returns a truth-value indicating whether the objects satisfy the relation. Equivalently, for every ordered set of *n* objects ($o_1...,o_n$) for which the relation holds, there is an explicit token of that relation: $R(o_1...,o_n)$. Any relation of this sort is referred to here as a unitary relation.

The distinction between a compositional and unitary representation is particularly relevant for concepts that are defined by a *relational structure*—the specific manner in which a concept's components are linked by their shared relations (Corral & Jones, 2014). Consider the statement *the dog chased the cat*. This statement has a different meaning than *the cat chased the dog* because of the specific manner in which the component parts in each statement are linked by the *chase* relation (*chase*(*dog, cat*) vs. *chase*(*cat, dog*)). Comprehending such statements seemingly requires representing them compositionally, based on the relationships among the statements' component parts.

Although compositional representations can be useful for distinguishing certain relational concepts, they can consume a large amount of working memory resources (Kintsch & Bowles, 2002), as the concepts' relations, components and their specific interconnections (i.e., role-filler bindings) must be explicitly represented. Representing a concept in this manner is potentially problematic because working memory resources are limited (Baddeley, 2003) and many relational concepts are fairly complex, making it challenging for people to explicitly represent

such concepts. In such cases, more economical representations might be more useful than representing a relational concept compositionally.

Although it has been acknowledged that it is possible for relational concepts to be represented in a non-compositional manner (Gentner, 1983, Footnote 4), theories of relational learning hold that people typically represent such concepts compositionally (Markman & Gentner, 2000). However, this assumption has not been directly examined under experimentally rigorous conditions. It thus remains unknown how people actually represent relational concepts. This project thus examines the conditions that might lead subjects to represent a relational concept compositionally and unitarily, with a specific focus on learning performance for different tasks, which might be facilitated by one of the two types of representations.

Representational Flexibility

One obstacle to addressing the issue of how people represent relational concepts is that people have a great deal of representational flexibility, such that many concepts can be represented in a variety of ways (Chalmers, French, Hofstadter, 1992; French, 1997; Mitchell & Hofstadter, 1990). For example, although a dam can be represented as an object that is used to store water or prevent floods, it can also be represented as a man-made object or as an engineering project that utilizes various scientific principles (e.g., gravity, conservation of energy) to produce, store, and release energy. This flexibility also seems to extend to how information for a given concept is represented. More specifically, the objects and relations of a stimulus can each be explicitly represented as individual units in working memory in a (1) non-structured manner, (2) in a structured manner, or (3) they might be chunked together into smaller units.

Many concepts contain featural and relational information, both of which are posited to be available for explicit representation (Gentner, 1983; Gentner & Markman, 1997; Markman, 1999; Markman & Dietrich, 2000; Markman & Gentner, 2000). For example, the concept of a dam contains information about attributes (e.g., large wall or barrier, flowing water) and relationships among those attributes (e.g., a barrier obstructs the flow of water). As mentioned above, theories of analogy and relational learning hold that *structured concepts* (i.e., concepts defined by a relational structure) are encoded, stored, and subsequently represented compositionally (Gentner & Markman, 1997; Markman, 1999; Markman & Dietrich, 2000; Markman & Gentner, 2000), such that people are explicitly aware of and attend to the components (i.e., attributes) and relations that define the given concept (Norman & Rumelhart, 1975; Schank & Abelson, 1977; Markman & Dietrich, 2000). The concept of a dam as an object that prevents floods, for example, can be compositionally represented as follows: *prevent*[(*obstruct*(*barrier*, *water*), *flooding*(*water*, *town*)].

There are important differences in the manner that featural and structured concepts are represented (Markman, 1999). A feature-based concept is defined by the presence of a given set of features, which can be represented in a vector space (e.g., featurally, the concept of car can be represented as {*doors*, *seats*, *windows*, *steering wheel*, *tires, engine…*}; Tversky, 1977). Although vector representations can be used to capture featural information, this type of representation may not be as useful for representing relational concepts. In order to represent relational concepts in a feature-based manner, the concept's objects and relations must be included as a collection of properties in a vector. However, such a representation lacks the information that defines a relational concept (i.e., the manner in which the concept's objects are linked together by their role-filler bindings). Consider a simple hypothetical scenario in which

Sam throws a pie at Robert. This scenario differs from one in which Robert throws a pie at Sam. Both scenarios consist of Sam, Robert, a pie that is thrown, an agent who plays the role of the thrower, and an agent who plays the role of the target. In the first scenario, Sam plays the role of the agent throwing the pie and Robert plays the role of the target; these roles are reversed in the second scenario. Using a feature-based representation, both scenarios would be represented as {*Sam*, *Robert, throws*, *pie*}. Even though such a representation contains all of the relevant objects and relations from both scenarios, it does not indicate how these elements are connected. Thus although the two scenarios differ, they cannot be distinguished from one another using a typical feature-based representation.

In contrast, this is not the case for representations that are structured, which can capture the relational complexity of a given scenario, and can therefore be used to differentiate between scenarios like the ones described above. Accordingly, it is has been argued that compositional representations are necessary in order to learn and represent structured concepts (Markman, 1999; Markman & Gentner, 2000). Markman and Gentner propose that although featural representations can be modified (e.g., representing a set of features configurally) to accurately reflect the structural information in a relational concept, doing so requires constructing a considerable number of representations, which can increase exponentially based on the concept's complexity. Consider a simple stimulus that consists of two side-by-side objects, each with two dimensions (size and brightness), in which the left object is bigger than the right and the right object is brighter than the left. This can be represented using a vector with configural features, such as [*left-object*, *right-object*, *bigger-object*, *smaller-object*, *brighter-object*, *darker-object*, *bigger-darker-right-object-left-object*, …]. This type of representation is problematic because

working memory is limited (Baddeley, 1986, 2003). Representing structured concepts through the use of configural features thus appears to be psychologically implausible.

Compositional Representations of Structured Concepts

The assumption that relational concepts are represented compositionally is critical to most models of analogical reasoning (Forbus, Gentner, Markman, & Ferguson, 1998; French, 2002; Gentner & Forbus, 2011; Gentner & Markman, 1997; Kokinov & French, 2003; Markman & Gentner, 2000; Morrison & Dietrich, 1995) and is arguably the foundational premise on which structure-mapping theory (Gentner, 1983) is based – the dominant theory of analogical learning and reasoning. The theory holds that two scenarios are only analogous if they share the same *relational structure*—the specific manner in which a system of objects and relations are linked by their role filler bindings (Corral & Jones, 2014). Consider a cup of water and a stadium filled with people: These two scenarios can be considered analogous because both share the same structure, such that each scenario's corresponding objects play the same roles and are linked by the same relation (i.e., cup and stadium fill the container role, and water and people fill the contained role); this structure can thus be represented as *contain*(*object*$_1$, *object*$_2$). According to structure-mapping theory, people discover analogies and relational concepts via comparison, whereby two scenarios and their corresponding objects and relations are put into alignment in a way that preserves the *parallel connectivity*—the corresponding components that are mapped between two scenarios are bound to the same role by the same relation(s)—between the elements in the two scenarios, leading to the abstraction of their common structure (Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983).

In line with this proposal, various studies have shown that people are sensitive to a concept's substructure (Corral & Jones, 2014; Goldstone & Medin, 1994; Goldstone, Medin, &

Gentner, 1991; Kotovsky & Gentner, 1996). Findings also indicate that similarity judgments between structured scenarios are determined by the extent to which they share a common structure (Clement & Gentner, 1991; Gentner & Kurtz, 2006; Markman & Gentner, 1993a, 1993b). Furthermore, people can accurately map objects from one scenario to its corresponding analogue in a different scenario (Keane, 1997; Markman, 1997; Spellman & Holyoak, 1992, 1996), even in cases where there are substantial superficial differences between the objects being mapped (e.g., bird-to-human; Markman & Gentner, 1993a). Markman and Gentner presented subjects with pairs of scenarios that contained cross mapped objects, in which the role that an object played in one scenario differed from its role in the other scenario; subjects were asked to map these objects to their corresponding analogues. The findings showed that object mappings were strongly driven by whether the objects played identical roles in each scenario, and not by their superficial similarities. Related work has shown that people are better able to recognize featural differences between scenarios that are connected to a common substructure (alignable differences) than differences that are not (i.e., non-alignable differences; Gentner & Markman, 1994), suggesting that discovering the shared structure between two scenarios draws attention to their different attributes (e.g., spaceships use rocket fuel and cars use gasoline; Markman & Gentner, 1993b). Taken together, these findings demonstrate that people can explicitly recognize and represent the elements and relational interconnections that define structured concepts.

However, it is not clear from these studies whether people typically represent structured concepts compositionally or default to other modes of representation. One possibility is that relationally structured concepts are only represented compositionally under certain circumstances, such as when people attempt to gain new knowledge or insight about a given scenario, and the concept's components must be used to make detailed predictions (e.g.,

scientific discovery). It is therefore possible that the artificial nature of these studies (e.g., asking subjects to explicitly map corresponding objects between scenarios) allowed for this uncommon phenomenon to be captured.

Representing Structured Concepts Unitarily

It is important to note that there are various types of structured concepts. For example, a structured concept can be defined by a *first-order relation*, which is a single relation that binds two objects (e.g., the left object is larger than the right object) or by a collection of relations among objects (e.g., the left object is bigger and brighter than the right object; Gentner & Markman, 2005). A structured concept can also be defined by a *higher-order structure*, which is a relational structure that contains relations among other relations (i.e., relations that take other relations as arguments), such as the greater mass of an object causes the object with less mass to revolve around it. In this scenario, the *cause* relation takes the *greater* and *revolves* relations as arguments (e.g., *cause*[(*greater*(*mass*(*Sun*), *mass*(*Venus*)), *revolves*(*Venus*, *Sun*))].

Various factors, such as prior knowledge about a concept's components and relations may affect how these properties are represented, which can change the structure of the concept. Consider a scenario where a baseball pitcher throws a ball that strikes a catcher's mitt. This scenario can be represented as a higher-order structure (i.e., *cause*[(*throw*(*pitcher*, *ball*), *strike*(*ball*, *catcher's-mitt*))]) or in a more compact manner, such as a three-place predicate, which is a ternary relationship among the pitcher, the ball, and the catcher's mitt (i.e., *strike*(*pitcher*, *ball*, *catcher's-mitt*)). Although the latter representation lacks the *cause* and *throw* relationships, they can be inferred by people with knowledge about the game of baseball, and hence do not need to be explicitly represented.

Evidence that people can make these types of inferences has been found by Rehder and Ross (2001). Rehder and Ross showed that prior knowledge about the relational interconnections among a concept's components can provide coherence to abstract concepts, which is posited to arise due to such knowledge binding the concept's relations, and thus imposing a coherent structure to the concept. This idea shares parallels with Murphy and Medin's (1985) proposal that peoples' naïve theories about the world bind the relations within a relational category together, thereby making such categories conceptually coherent. Rehder and Ross constructed stimuli that consisted of three short sentences that described different components of machines that work to collect or remove waste products; half of the stimuli satisfied relationships (believed to be known by subjects but not explicitly present in the stimuli) among the machine's components, whereas the other half did not. A stimulus could therefore be classified as a member of either a coherent or incoherent category. Consider the following description of a coherent stimulus: operates in war zones, works to gather shards of metal, and has a large magnet. People are presumed to know that shards of metal can be found in war zones and can be collected by large magnets. The coherence in the stimulus thus arises due to subjects' a priori knowledge about these interconnections. In contrast, consider an example of an incoherent stimulus: operates in warzones, works to remove carbon monoxide, and has a finely woven net. Carbon monoxide is not typically associated with warzones and it cannot be removed by a finely woven net. Rehder and Ross found that subjects were better able to learn the category rule for coherent items than for items that were incoherent. Because the only difference between coherent and incoherent items was subjects' presumed knowledge about the manner in which the concepts' components were interconnected, these findings suggest that such knowledge can change the manner in which a relational structure is learned and represented.

Corral et al. (2017) raise the possibility that such a priori knowledge may allow for a concept to be represented unitarily, in which a relationally structured stimulus is encoded and represented as an atomic property, devoid of its substructure, that can be recognized directly in the stimulus. In such cases, a priori knowledge gives the unitary representation meaning, by providing the background knowledge needed for unpacking the unitary attribute. Consequently, it may be possible to process such representations in a manner that is psychologically similar to features. For example, a coherent stimulus from the Rehder and Ross (2001) experiments might be represented unitarily as an object that functions (e.g., *functions*(*object*)), whereas an incoherent stimulus might be represented as an object that does not function. One possibility is that people become aware of a stimulus' unitary properties when its components and relations activate higher-order concepts in long-term memory (LTM), which are subsequently ascribed to the stimulus and used to determine its category membership (e.g., the stimulus functions and therefore belongs to a category that is defined by objects that function). Thus, structured concepts that can be defined by general themes or ideas that have a corresponding representation in LTM (e.g., the story is about betrayal) might be more readily represented unitarily than other concepts.

Unitary representations may be formed through an automatic chunking mechanism, in which the components and relations of a structured stimulus or scenario are automatically grouped to form a conceptual unit. Automatic chunking appears to be ubiquitous in daily functioning, as people routinely combine the perceptual components of objects into meaningful representations (Czerwinski, Lightfoot, & Shiffrin 1992; Gobet et al., 2001; Goldstone, 2000; Shiffrin & Lightfoot, 1997), without necessarily attending to its component parts. For example, although a car consists of various components (e.g., tires, seats, a steering wheel), most people

can immediately recognize one without attending to or being aware of each of its parts. A similar process occurs during reading, in which individual features are chunked together to form semantically meaningful representations (e.g., characters are chunked to represent words and words are chunked to represent sentences; Healy, 1976; Simon, 1974). Chunking allows people to bypass their working-memory limitations and process a greater amount of information than would otherwise be possible (Ericcson, Chase, & Faloon, 1980; Miller, 1956).

In theoretically related work, Clark (2006) posits that tagging or labeling a given concept allows for the concept to be tokenized, such that a tag or label can be used to represent the concept as a whole, and may thus facilitate the process of chunking conceptual information. For example, consider an instance in which two objects share a given relationship with one another on at least one dimension. Although this description can be used to describe such an instance, the concept of *similarity* can be invoked to represent the same information (i.e., the two scenarios are similar). As this example illustrates, tags and labels can be used to represent higher-order concepts in a lower-order or feature-based manner. Indeed, labels can often aid people in acquiring complex category rules (Lupyan, Rakison, & Mclelland, 2007). Furthermore, research involving non-human primates has shown that the use of tags and labels allows chimpanzees to learn relational concepts (e.g., *greater than*, *sameness*) that they are otherwise unable to learn (Boysen, Bernston, Hannan, & Cacioppo, 1996; Thompson, Oden, Boysen, 1997). To state this argument more specifically, it is possible that non-humans can acquire concepts that are defined by single relations, but not those that comprise a higher-order structure. However, replacing a first-order relation with a token might enable that token to become an argument to another first-order relation, which can further facilitate the learning of more relationally complex concepts through the iterative use of this chunking process. Thus, tags enable higher-order relations to be

acquired and encoded using learning mechanisms that do not rely on structured representations. It has been argued that non-human animals lack the cognitive machinery to acquire relational concepts in a compositional manner, and thus likely rely on more primitive forms of learning (e.g., feature-based) that bypass the explicit representation of a concept's components and relations (Penn, Holyoak, & Povinelli, 2008). If Penn et al. are correct, studies such as Boysen et al. and Thompson et al. demonstrate that relational concepts can indeed be represented and acquired in a non-compositional manner, and moreover that the use of tags and labels appear to facilitate such processing.

The idea that relational concepts are represented non-compositionally can also be found in various influential models of metaphor comprehension (e.g., Glucksberg & Keysar, 1990; Ortony, 1979). One such model is the attributive categorization model (Glucksberg, 2003), which holds that metaphors are understood as statements that indicate the base and target are members of the same category, which is defined by a shared higher-order attribute. According to this model, statements such as the *sprinter* (target) *was faster than a speeding bullet* (base) are understood to be assertions about the shared category membership between the sprinter and a speeding bullet, such that both pertain to a category that is defined by objects that travel at a relatively high velocity. Because category membership is determined by a given attribute (e.g., objects that travel at a high velocity), it can therefore be recognized directly in the stimuli as a shared global property of the base and target, obviating the need for alignment (Kintsch & Bowles, 2002). Importantly, this model runs counter to the structure-mapping account of metaphor comprehension, which posits that metaphors are understood by mapping the corresponding components and relations between the base and the target (Gentner & Bowdle, 2008; Gentner & Wolff, 1997).

Related work has shown that there is an asymmetry in metaphor comprehension that depends on which concept is used as the base and which is used as the target, a prediction that follows directly from the attributive categorization model (Glucksberg, McGlone, & Manfredi, 1997). The statement *the sprinter was faster than a speeding bullet* may therefore be understood to hold a different meaning than *the speeding bullet was faster than the sprinter*. One reason for this asymmetry might be due to people basing their inferences about the target on the attribute that is most salient in the base, which often differs among concepts. For example, precision might be a salient attribute that comes to mind when people think of surgeons, whereas creativity might be salient when they think about artists. Different inferences can hence be made about the target for the statement *the artist is a surgeon* (i.e., the artist is precise) than for the statement *the surgeon is an artist* (i.e., the surgeon is creative).

The argument can be made that because the components and relations that are analogous between the base and target are the same regardless of which item is used as the base and which is used as the target, a structure-mapping framework cannot account for the asymmetry in metaphor comprehension found by Glucksberg et al. (1997). However, because the concept that is used as the base can affect which dimensions in the target are most salient and each dimension can consist of different components and relations, the elements that are mapped between analogous concepts may differ depending on which concept is used as the base. This is a critical point because the base and target may not be alignable on many of their dimensions. A structure-mapping account of metaphor comprehension therefore seems to predict that comprehension should decrease in cases where the salient dimensions in the base are not analogous or cannot be fully aligned with the salient dimensions in the target.

Another prediction that seems to follow from structure-mapping theory is that comprehension should be faster for literal statements than those that are figurative, as literal statements do not appear to require a mapping process, whereas this is not the case for figurative statements. However, various studies have failed to show support for this idea and have found no differences in the time it takes subjects to comprehend the meaning of literal and figurative statements (Giora, 1999; Harris, 1976; McElree & Nordlie, 1999; Tartter, Gomes, Dubrovsky, Molholm, & Stewart, 2002). Furthermore, it appears metaphors are often understood automatically and seem to require a minimal amount of explicit processing (Glucksberg, 2003; Gildea & Glucksberg, 1983; Glucksberg, Gildea, Bookin, 1982; cf. Lai, Curran, & Menn, 2009). Taken together, these findings run counter to a structure-mapping account of metaphor comprehension and the idea that such concepts are represented compositionally, as the mapping of the concepts' corresponding components and relations is posited to be an explicit, arduous process that consumes a large amount of working memory resources (Kintsch & Bowles, 2002). In contrast, these are the exact pattern of results one would expect if structured concepts were represented unitarily, as the concept can be treated as a feature and recognized directly in the stimulus in much the same way that people can immediately recognize the color or shape of an object.

Anecdotal evidence for this phenomenon can be found in language, in which people routinely use words that consist of an intricate substructure that may not be explicitly represented. Consider the concept of *aid*: Aid consists of at least two agents or objects, at least one of which has a goal, at least one obstacle that impedes that goal, an action or outcome produced by an agent that removes or reduces the impact of the obstacle, and a specific relational pattern that binds these components in the appropriate manner (e.g., Jill provided the short-

staffed homeless shelter aid by volunteering). Although these relations and components can be explicitly generated, they may not be stored or explicitly represented when people think of the concept of aid.

Moreover, if relational concepts are indeed represented compositionally (Markman & Gentner, 2000), people should be able to explicitly describe a concept's components and their relational interconnections; however, this is often not the case. Although people can learn various complex relational concepts, they might often fail to encode and store those concepts' components and relations (Keil, 2003a). Encoding the relational components of a structured concept requires a large amount working memory resources (Kintsch & Bowles, 2002), which may lead to such information being neglected, as people may instead rely on unitary properties that can adequately capture the general gist of the concept. These properties may serve as placeholders for the stimulus' substructure that is unknown.

The same may hold true for a variety of complex relational scenarios we encounter on a daily basis (Keil, 2003a, 2003b, 2006). Although people can be quite confident in their understanding of various types of phenomena (e.g., the concept of gravity), when prompted to explain such phenomena, accounts are often incomplete and lack explanatory depth (Rozenblit & Keil, 2002). Furthermore, these explanations typically contain scarce references to the components and relational interconnections that define a given phenomenon. In these instances, subjects are surprised by their inability to describe phenomena that they previously expressed high confidence in understanding. These findings suggest that the compositional elements of various types of relational phenomena are either not encoded and stored or cannot be accessed, and thus may not play a prominent role in the manner that such concepts are explicitly represented and understood. Furthermore, this possibility may account for the reason that

learning and comprehension are improved when people are asked to explain a given phenomenon (Chi, de Leeuw, Chiu, & LaVancher, 1994), as this process may allow people to logically construct and explicitly represent and access an item's components that were previously not known.

Keil (2003a, 2003b, 2006) posits that people form skeletal representations of the substructure that defines relational phenomena, which only contains a small subset of the concept's component parts. Nevertheless, such representations can facilitate a general understanding of relational phenomena (e.g., its outcome or function), which allows people to vaguely explain and accurately predict the outcomes of events that correspond to a given phenomenon.[1] For example, a layperson who is asked to explain the concept of gravity might describe it as an anchoring force that keeps objects grounded or that causes solid objects to fall when they are not resting on a solid surface. Although such accounts lack critical information, they exemplify a general understanding of the concept of gravity that can be used to effectively navigate the environment. Such theme-like descriptions (which are mostly devoid of references to a substructure; Rozenblit & Keil, 2002) seem well suited to be represented as unitary concepts (e.g., *anchor*(*gravity*)), and moreover appear to more closely resemble the descriptions that would be expected for concepts that are represented unitarily (i.e., descriptions with minimal reference to structured components) than for those that are represented compositionally.

Nevertheless, this is not to argue that a concept's components and relations do not play an integral role in relational learning. One of the primary differences between experts and novices resides in the manner that experts represent concepts within their domain (Chi, Feltovich, & Glaser, 1981). Chi et al. found that experts identify such concepts by their substructure and

---

[1] This proposal shares similarities with Dennett's (1987) theory on intentional stance.

display rich knowledge about these concepts' components and their relational interconnections, indicating that compositional knowledge can indeed be accessed and explicitly represented. In contrast, novices typically neglect and fail to fully learn a concept's relational structure and instead attend to their surface features. There is also evidence that experts can readily chunk large relational systems into individual units based on overlapping conceptual themes (e.g., the configuration of chess pieces represents a *fork*—an instance in which at least one chess piece can simultaneously attack multiple opposing pieces), which can greatly enhance memory of those concepts and their interconnections among its components (i.e., relational structure; Chase & Simon, 1973). Taken together, these findings suggest that unlike novices, experts can access the unitary and compositional properties that define the concepts within their domains. Hence, both types of representations likely play a prominent role in the development of insight and concept discovery.

However, it is unclear which type of representation, compositional or unitary, develops first or how the two might coevolve with one another. The clearest evidence that relational concepts are represented compositionally comes from studies on expertise (e.g., Chi et al, 1981), which takes people many years to develop (Ericsson, Krampe, & Tesch-Römer, 1993), suggesting that such representations require an extensive amount of time to materialize and may often precede the formation of unitary representations. One obstacle to understanding the nature of knowledge representation is that the tasks that are used to make such assessments might affect and transform the initial representation and may therefore not reflect the role that such knowledge plays in analogical reasoning. For instance, when explaining a given phenomenon, subjects may discover and construct its relational structure on the fly, which may not have been explicitly known or represented beforehand. Such an occurrence might lead researchers to draw

erroneous conclusions, both about the manner in which such knowledge is typically represented and how those representations are used in analogical reasoning.

Overview

One issue that many studies on relational learning face is that they typically use complex training paradigms that consist of various components. These paradigms include (1) the manner in which similarity between two analogous concepts is described to subjects, (2) asking subjects to explicitly compare two items, (3) asking subjects to write out similarities and/or differences between those items, and (4) an inference-based reasoning task. Each of these components seems to encourage subjects to represent the stimuli compositionally, as they draw attention to the stimuli's component parts. It is therefore unclear how each of these factors affects how a relational concept comes to be represented or how such concepts might best be learned. The purpose of this paper is to tease apart these factors to better understand the relationship between relational learning and concept representation. This paper reports seven experiments that examine how people typically represent relational concepts (Experiment 1), how such representations affect subsequent relational learning (Experiments 2-4), and the factors that affect concept representation and learning (e.g., different types of comparisons and tasks; Experiments 5-7).

CHAPTER II

Experiment 1

This study examines whether subjects typically represent relational concepts unitarily or compositionally. Common relational nouns might be particularly useful for investigating this question, as these concepts have rich relational structures that subjects are likely to be familiar with (due to the commonality of these concepts). Furthermore, due to this familiarity, subjects might be able to readily represent relational nouns unitarily (if relational concepts can indeed be represented in such a manner). Thus, one benefit of using common relational nouns as stimuli is that subjects might be capable of representing such concepts both unitarily and compositionally, and might therefore be able to readily access both. For this reason, such concepts might be well-suited for testing which type of representation subjects prefer.

The stimuli in this study were thus made up of common relational nouns (e.g., *tradeoff*, *aid*), each of which consisted of two definitions, which were both correct. For each noun, one definition consisted of a unitary-based description (i.e., a synonym) and the other consisted of a compositional-based description. For example, the unitary definition for *tradeoff* was "*A situation in which a compromise or concession is made*", whereas the compositional definition was "*A situation in which an agent must choose between or balance two or more things that are opposite or cannot be had at the same time*". Figure 1 shows an example stimulus from this study. Subjects were asked to select the definition that best represented how they typically think about the corresponding noun. Table 1 contains all of the relational nouns and their corresponding definitions that were used in this study.

Please select the option that best captures how you typically think of **aid**:

○ An instance in which an agent's actions make it easier for another agent to reach their goal
○ An instance in which help or assistance is provided

Figure 1. An example stimulus for Experiment 1 for the concept of *aid,* along with its compositional (option *a*) and unitary definitions (option *b*).

One concern worth noting is that the unitary definitions are shorter in word length and contain less information than the definitions that are described compositionally. Although steps can be taken to remedy this issue, doing so carries the risk of making the unitary definitions artificial and incoherent to subjects, which might affect the definitions they select. Because the primary goal of this study is to examine subjects' typical preference between unitary and compositional concepts, a tradeoff must be made between experimental control and ecological validity, with a greater emphasis on the latter for the purposes of this study. In line with this goal, the differences between the two types of descriptions embody the proposed characteristics of compositional and unitary concepts, such that unitary concepts contain less explicit information than concepts that are represented compositionally (described in detail above).

For each noun, all subjects were presented two definitions and were asked to select which best captures how they typically think about the noun. For conciseness, these questions will be referred to as representation questions; the main task only consisted of representation questions. A control group was used to provide a baseline measure of how relational nouns are typically represented, and thus these subjects were only presented representation questions on the main task. Previous work posits that relational concepts are represented compositionally (Markman & Gentner, 2000), and thus one prediction is that subjects in the control condition will select a higher proportion of compositional definitions than unitary definitions. However, one possibility

is that because unitary representations are posited to be more computationally efficient (as discussed above; Corral et al., 2017), subjects in the control condition will select the unitary definitions at a higher proportion than those that are described compositionally.

Table 1. List of nouns and each of their corresponding compositional and unitary definitions that were used in Experiment 1.

Concept	Compositional Definition	Unitary Definition
Tradeoff	A situation in which an agent must choose between or balance two or more things that are opposite or cannot be had at the same time	A situation in which a compromise or concession is made
Aid	An instance in which an agent's actions make it easier for another agent to reach their goal	An instance in which help or assistance is provided
Reciprocity	A situation in which two or more agents or groups agree to do something similar for each other, to allow one another shared benefits	A mutual trade or exchange
Showdown	An agent, object, or circumstance that impedes the passage or progress of another agent, group, or object	A confrontation or clash
Obstruction	An agent, object, or circumstance that impedes the passage or progress of another agent, group, or object	An obstacle, barrier or barricade, a blockage or hindrance
Risk	A situation or action that exposes an agent or group to the possibility of harm or an undesired outcome	A situation in which there is potential danger
Investigation	An inspection carried out by an agent who's primary goal is to find information about a given question	To search or examine
Cooperation	An instance in which two or more agents or groups have at least one goal in common and work together to accomplish that goal	A partnership or collaboration
Compensation	An act that is intended to make up for or balance the loss that an agent or group has incurred, typically due to the fault of another agent or group	An act of reparation or repayment
Deception	An instance in which one or more agents provide inaccurate information to another agent or group with the purpose of misleading them	To lie or deceit, to be dishonest

In addition to the control group, three experimental conditions (mapping, definitional, and generation) were used to examine which factors might change how subjects typically represent relational concepts. For these conditions, the corresponding manipulation was presented on each trial before subjects were presented a representation question. Subjects in the mapping condition were presented with two short scenarios (as shown in Figure 2) about a given noun and were asked to map the parts among the scenarios that were analogous (as shown in Figure 3); subjects in the definitional condition were asked to write out a definition for the noun in each trial; and subjects in the generation condition were asked to write out a short scenario about the noun for the given trial (e.g., in 3-4 short sentences, make up a scenario in which a tradeoff must be made).

Please read each scenario carefully.

Aid

Scenario 1	Scenario 2
Robert broke his leg in a boating accident 1 week ago and it is still difficult for him to get around the house. Robert has mostly been staying at home with his trusty dog Max. Robert is watching television on his couch and wants to change the channel, but left the remote control on his coffee table. It is still very painful for Robert to stand once he has been sitting so he does not want to walk over to the coffee table. Max runs to the coffee table and brings the remote control to Robert. Robert is now able change the channel without having to get up off the couch.	Tom is enrolled in a calculus course and has done poorly on the previous 2 exams and needs to score an 82 on the final to pass the course. However, Tom does not understand some of the material that was covered in the past 7 weeks. Marcy, the top student in the course, recognizes that Tom is struggling and offers to tutor him the week before the final. For one week, Marcy tutors Tom for 3 hours a day. Marcy is able to explain the material in a way that makes sense to Tom. Tom ends up scoring a 97 on the final and passes the course.

Figure 2. The two scenarios that instantiate the concept of *aid* in the mapping condition in Experiment 1.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

Scenario 1

A. Max runs to the coffee table and brings the remote control to Robert.

B. Robert is now able change the channel without having to get up off the couch.

C. Robert is watching television on his couch and wants to change the channel, but left the remote control on his coffee table. It is still very painful for Robert to stand once he has been sitting so he does not want to walk over to the coffee table.

Scenario 2

D. Tom is enrolled in a calculus course and has done poorly on the previous 2 exams and needs to score an 82 on the final to pass the course.

E. Marcy, the top student in the course, recognizes that Tom is struggling and offers to tutor him the week before the final.

F. Marcy is able to explain the material in a way that makes sense to Tom. Tom ends up scoring a 97 on the final and passes the course.

Passage A ▢
Passage B ▢
Passage C ▢

Figure 3. The mapping portion of the trial in the mapping condition for the concept of *aid* in Experiment 1.

All three of these manipulations were intended to draw subjects' attention to the concepts' component parts. More specifically, the two scenarios for each relational noun in the mapping condition emphasize the corresponding concept's relational structure, and the mapping process is intended to further highlight the relational interconnections among the concept's components. The definitional condition required subjects to explicitly describe each relational noun, which was intended to make the concept's relational structure more salient to subjects; for each noun, subjects were only instructed to write out a definition so as not to explicitly bias them towards a given representation. Likewise, the process of generating a scenario about each relational noun might force subjects to explicitly represent each of the concept's component parts. Therefore, one prediction that follows is that subjects in the experimental conditions will

select the compositional definitions at a higher proportion than subjects in the control condition. Alternatively, one possibility is that writing out a definition or generating a scenario about a given concept will simply reflect how a subject represents that concept, and thus the compositionality in subjects' definitions and scenarios might predict the proportion of compositional definitions that they select on the representation questions.

Lastly, this study also examines the relationship between relational concept representation and other cognitive attributes. More specifically, at the end of the study all subjects completed a modified version of the forward digit span, and a revised version of the Need for Cognition Assessment (Cacioppo, Petty, & Kao, 1984), which respectively provide a measure of subjects' verbal working memory (Richardson, 2007) and the gratification they derive from engaging in reflective thought (Cacioppo & Petty, 1982). Because compositional representations are computationally expensive (Forbus, Gentner, & Law, 1995), one possibility is that subjects who score higher on the forward digit span task have a greater amount of cognitive resources to work with when representing relational concepts, and are thus more likely to represent such concepts compositionally than subjects who score lower on the forward digit span task. Alternatively, it is possible that subjects who score higher on the forward digit span task, because they might have a greater amount of cognitive resources to work with than subjects who score lower on this task, are more likely to consolidate a relational concept into a computationally efficient representation, and hence represent such concepts unitarily. A similar set of predictions can be made for subjects who score higher on the Need for Cognition Assessment, as subjects who enjoy engaging in reflective thought might be more likely (than subjects who do not enjoy this process) to prefer the more complex representation that takes more work to discover, and therefore represent relational concepts compositionally.

Alternatively, these subjects might devote an extensive amount of time to thinking about the structure of a given relational concept (because they enjoy engaging in reflective thought) and come to recognize that the concept can be consolidated into a more compact representation. As a result, these subjects might prefer to represent such concepts unitarily.

Method

Participants

One hundred sixty-one subjects were paid $1.25 for their participation in this study, which was conducted online through Mechanical Turk. Only subjects who had a 95% hit approval rating and who had completed over 500 Mechanical Turk studies were allowed to sign up for this study. Subjects were randomly assigned to four conditions (between-subjects): control ($N = 48$), definitional ($N = 44$), generation ($N = 33$), and mapping ($N = 36$).

Design and Materials

All stimuli were presented on a computer monitor on a white background. All subject responses were entered using a computer keyboard and a computer mouse or touchpad. Subjects in the definitional and generation conditions were provided a textbox to type in the definitions and scenarios that they were asked to provide. The stimuli consisted of 10 relational nouns (see Table 1). For the compositional definitions, the standard definition of each noun was modified to highlight the elements of the noun's relational structure. For the unitary definitions, synonyms were used to define each noun; these definitions did not contain relationally structured information. For the mapping condition, two analogous scenarios were created for each noun, each of which instantiated the noun's relational structure. Appendix A contains all of the

scenarios that were used in the mapping condition, along with the corresponding elements that subjects were asked to map.

Table 2. Revised (18-item) Need for Cognition Scale. This table was recreated from Cacioppo et al. (1984).

Item Number	Item Wording
1.	I would prefer complex to simple problems.
2.	I like to have the responsibility of handling a situation that requires a lot thinking.
3.	Thinking is not my idea of fun.*
4.	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.*
5.	I try to anticipate and avoid situations where there is a likely chance I will have to think in depth about something.*
6.	I find satisfaction in deliberating hard and for long hours.
7.	I only think as hard as I have to.*
8.	I prefer to think about small, daily projects to long-term ones.*
9.	I like tasks that require little thought once I've learned them.*
10.	The idea of relying on thought to make my way to the top appeals to me.
11.	I really enjoy a task that involves coming up with new solutions to problems.
12.	Learning new ways to think doesn't excite me very much.*
13.	I prefer my life to be filled with puzzles that I must solve.
14.	The notion of thinking abstractly is appealing to me.
15.	I would prefer a task that intellectual, difficult and important to one that is somewhat important but does not require much thought.
16.	I feel relief rather than satisfaction after completing a task that required a lot of mental effort.*
17.	It's enough for me that something gets the job done; I don't care how or why it works.*
18.	I usually end up deliberating about issues even when they do not affect me personally.

Note: * This item was reverse scored.

A modified version of the forward digit span task was given to all subjects. The sequence of digits that subjects were asked to recall ranged from 3-11, starting with 3 on the first trial and increasing by one digit on each subsequent trial. If a subject recalled the sequence of digits for a

given trial correctly, they continued on to the next trial, otherwise they did not. The task thus ended once a subject incorrectly recalled the sequence of digits that were presented on a single trial or after the 9th trial (i.e., the task contained a minimum of one trial and a maximum of nine).

The Need for Cognition Assessment consisted of 18 items, which were presented in the order used by Cacioppo et al. (1984). Subjects entered their responses using a Likert scale, which ranged from 1-5 (extremely uncharacteristic of me to extremely characteristic of me), by clicking on the option of their choice. Table 2 contains the Need for Cognition items, along with the order in which they were presented.

Procedure

At the start of the study, all subjects were notified that they would be shown various words, each of which would consist of two definitions, and that they should select the definition that best reflects how they typically think about the word. Subjects were also told that for each word, both definitions were correct and that there were no right or wrong answers. Subjects in the mapping condition were told that before selecting a definition for the word on the given trial, they would be asked to read two side-by-side short scenarios about the word and would then need to match the parts between the two scenarios that were most similar to one another. Subjects in the definitional condition were told that on each trial, they would first be asked to define the word for that trial and then select the definition that best reflects how they typically think about that word. Subjects in the generation condition were told that on each trial, they would first be asked to create a short scenario about the word for that trial and then select the definition that best reflects how they typically think about that word. At the end of the instructions, all subjects were asked to click on the "next" button, which was located on the bottom right side of the screen, to proceed.

The order in which each noun was presented was randomized for all subjects, as was the spatial position in which each definition was presented on the screen. For all subjects, a bubble was presented directly to the left of each definition (as shown in Figure 1), and subjects selected a given definition by clicking on the corresponding bubble. After selecting a definition, subjects were allowed to move on to the next trial by clicking on the "next" button, which was located on the bottom right side of the screen. Subjects were not allowed to move on to the next trial without selecting a definition and were not allowed to navigate to a previous trial.

For the main task, subjects in the control condition were only presented representation questions (i.e., subjects were asked to select the definition for each word that best represented how they typically thought about the word). Each trial in the definitional and generation conditions consisted of two phases. In the definitional condition, subjects were asked to write out a definition for each noun; in the generation condition they were asked to write out a short scenario (in 3-4 sentences) about the noun. For both of these conditions, a representation question was presented to subjects in the second phase of the trial. To move on to the next phase of the trial or to move on to the next trial, subjects in all conditions were required to click on the "next" button, which was located on the bottom right side of the screen.

Each trial in the mapping condition consisted of three phases. In the first phase, subjects were presented two side-by-side short scenarios about the corresponding noun (as shown in Figure 2) and were asked to carefully read each. The scenarios on the left and right were labeled "Scenario 1" and "Scenario 2", respectively. These labels were presented directly above the corresponding scenario, as shown in Figure 2. For each noun, the scenarios that were presented on the left and right side of the screen were randomized before the study, and were thus in a fixed order and the same for all subjects. In the second phase, subjects were presented three

passages from each scenario, which emphasized a different element of the noun's relational structure. For example, in Figure 3, Passage C from Scenario 1 and Passage D from Scenario 2 both instantiate the same elements of the relational structure for *aid*, such that in both cases there is an agent who has a goal and there is an obstacle that is obstructing that goal. Additionally, there was a one-to-one mapping between the passages from the two scenarios, such that only one passage from Scenario 1 could be mapped to a passage from Scenario 2. The passages for Scenario 1 were labeled A-C and were presented on the left side of the screen and the passages for Scenario 2 were labeled D-F and were presented on the right side of the screen (see Figure 3). The order in which the passages were presented on the screen was randomized for each subject. On the bottom left side of the screen, subjects were presented three labels (Passage A, Passage B, Passage C), each of which had a textbox next to it (as shown in Figure 3). Subjects were asked to indicate which passages from the two scenarios were most similar to one another by typing in the passage's letter option from Scenario 2 into the textbox next to the corresponding passage label for Scenario 1 (as shown in figure 3). The third phase of the trial consisted of a representation question.

After completing the main task, subjects were given a modified version of the forward digit span task, wherein subjects were shown a sequence of digits and were asked to type them into a textbox in the order that the digits were presented. This task ranged from 1-9 trials, depending on when a subject incorrectly recalled the sequence of digits that were presented in a given trial. Each digit was presented in bold font, one at a time, and was shown at the center of the screen for 1 second. Prior to the study, a random sequence of digits was generated for each trial. These sequences of digits were used for all subjects. On each trial, after the last digit was shown on the screen, a textbox was presented directly below where the digits were shown and

subjects were asked to type in the sequence of digits for that trial. Lastly, subjects were asked to complete the Need for Cognition Assessment. On both of these tasks, after entering a response, subjects were able to move on to the next trial or item by clicking on the "next" button, which was located at the bottom right side of the screen. Subjects were not allowed to move to the next trial or item without entering a response and were not allowed to return to a previous trial or item. The study lasted an average of 23 minutes.

Results and Discussion

First, an analysis was conducted to examine if subjects selected one type of definition at a greater proportion than the other. The proportion of compositional definitions that were selected by each subject was recorded. Figure 4 shows the proportion of compositional definitions that subjects in each condition selected. A one-sample t-test was conducted, which examined whether this proportion was different from the proportion of compositional definitions that subjects would be expected to select if they had no preference for either type of definition (i.e., 50%). This analysis revealed that subjects selected a lower proportion of compositional definitions (M = .374, SE = .018) than would be expected if they had no preference for either type of definition, $t(160)$ = 6.73, p < .0001. This finding suggests that subjects had a strong preference for the unitary definitions over the definitions that were described compositionally.
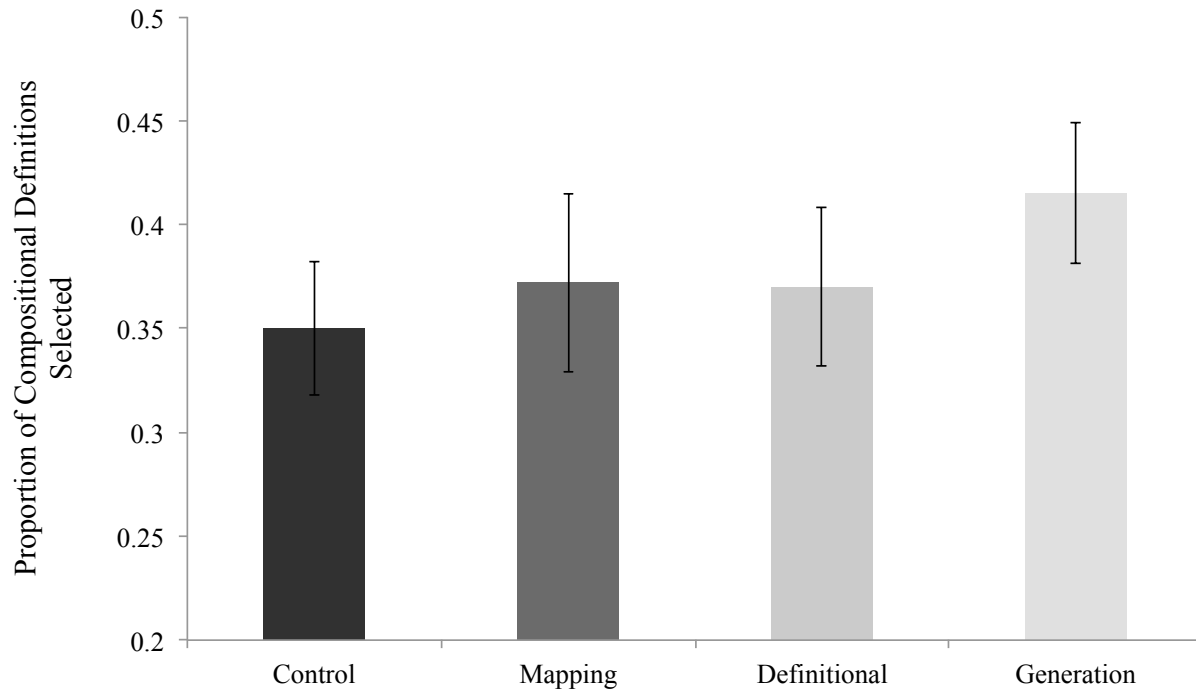
Figure 4. Shows the proportion of compositional definitions that subjects in each condition, along with the standard error of the mean in Experiment 1.

Next, a one-way ANCOVA was conducted to examine whether there were differences among the groups in the proportion of compositional definitions that subjects selected; scores on the Need for Cognition Assessment and the forward digit span task were included in the model as covariates. However, no differences were found among the groups ($M_{control}$ = .350; $M_{mapping}$ = .372; $M_{definitional}$ = .371; $M_{generation}$ = .415) in the proportion of compositional definitions that were selected, $F(3, 155) = .50$, $p = .68$, $MSE = .056$. Performance on the Need for Cognition Assessment ($M = 54.86$, $SD = 3.52$) was also not a reliable predictor of the proportion of compositional definitions that subjects selected, $F(1, 155) = 1.24$, $p = .267$, $MSE = .056$. However, there was a marginal, negative relationship between performance on the forward digit span task ($M = 6.11$, $SD = 2.39$) and the proportion of compositional definitions that subjects

selected, β = -.15, *F*(1, 155) = 3.68, *p* = .057, *MSE* = .056, such that subjects who had a lower digit span score were more likely to select a higher proportion of compositional definitions than subjects who had a higher digit span score. Figure 5 shows the relationship between subjects' digit span score and the proportion of compositional definitions that they selected.
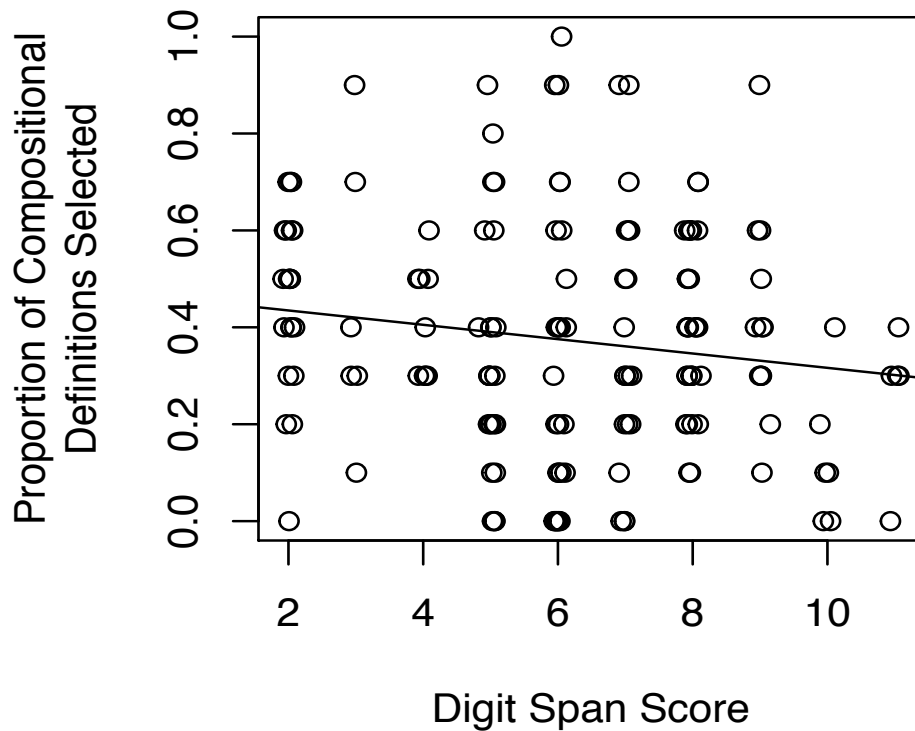
Figure 5. Illustrates the relationship between subjects' digit span score and the proportion of compositional definitions that they selected on representation questions in Experiment 1, with the line of best fit.

Additionally, two post hoc tests were conducted. The first test examined whether subjects in the control group differed in the proportion of compositional definitions they selected from subjects in the generation condition; the generation condition was chosen for this comparison because these subjects selected the highest proportion of compositional definitions from the three

experimental conditions. Nevertheless, only a non-significant trend was observed, $t(79) = 1.34$, $p = .185$, $d = .306$. The second post hoc test combined all of the three experimental conditions and compared the proportion of compositional definitions that these subjects selected to that of the subjects from the control group, but no differences were found, $t(159) = .83$, $p = .41$.

Taken together, these findings suggest that although it is plausible that subjects prefer to represent relational nouns unitarily, these representations might not be particularly mutable, as none of the experimental conditions produced an observable change in the type of definitions that subjects selected. However, it is worth noting that a non-significant trend was observed between the generation and control condition, as subjects in the generation condition selected a greater proportion of compositional definitions (by 6.5%) than subjects in the control condition. Although this difference was not statistically reliable, the qualitative difference between the two groups is encouraging and points to a way in which subjects' mode of representation might be altered. More specifically, creating a scenario about a given concept (as subjects in the generation condition were asked to do) might explicitly shift subjects' attention to the concepts' component parts. Thus, similar manipulations might prove to be useful in shifting how subjects represent relational concepts.

One explanation for the present findings is that the manipulations that were used were not strong enough to change how subjects' typically represent relational nouns. Strengthening these manipulations might therefore produce the intended change in subjects' representations. One way to accomplish this goal might be to combine and modify some of the manipulations used in this study into a single condition. For instance, it is feasible to combine and adapt the generation and mapping conditions, wherein for each relational noun subjects are first asked to write out a short scenario about the noun and are then shown two analogous scenarios about that noun and

are asked to align them; or as an alternative, subjects could be asked to map the scenario that they generate to an analogous scenario. To strengthen this type of manipulation, after mapping the corresponding passages, subjects could be shown the correct responses and then be asked to write out a short explanation of why each paired passage is analogous. This additional process would require subjects to further think about the concept's relational structure, and might therefore help to modify how subjects typically represent relational nouns.

It is also worth noting that a different result might be observed with different types of relational concepts. One potential issue that might have worked against the manipulations producing a change in how the stimuli are represented is that the relational nouns that were used are fairly common and thus subjects were likely highly familiar with them. As a result, these concepts might have been unitized long ago and their representations might not be particularly amenable to change. A different result might therefore be observed with relational concepts that subjects have less familiarity with, such as scientific theories. Furthermore, subjects might also be less likely to represent such concepts unitarily and might hence be more likely to select a higher proposition of compositional definitions for relational concepts that are less common than for those that are highly familiar.

One possibility to consider is that there are individual differences in how subjects represent relational nouns, such that some subjects represent these concepts compositionally and others represent them unitarily. Critically, subjects' written responses in the definitional and generation conditions might provide a measure of how each subject typically represents the corresponding concept, which might in turn predict the type of definition that subjects select on the representation questions. To examine this possibility, a regression was conducted which used subjects' written responses in the definitional and generation conditions to predict the type of

definitions that they selected.[2] In order to conduct these analyses, the definitions and scenarios that subjects wrote out were scored (by the researcher) based on the compositionality in their responses (from 0-2) and an average compositionality score was computed for each of these subjects. Responses that made no explicit reference to the relational structure of the corresponding noun were given a score of 0; responses that made some reference to the corresponding noun's relational structure were scored as a 1; and responses that fully captured the noun's relational structure were scored as a 2. Figure 6 shows the relationship between the proportion of compositional definitions that were selected and the average compositionality of subjects' scores, by condition.

As hypothesized, the results showed that subjects' average compositionality scores reliably predicted the proportion of compositional responses that they selected, $\beta = .495$, $t(75) = 4.93$, $p < .0001$, such that subjects who had a higher compositionality score selected a higher proportion of compositional definitions. Additionally, this result held when a regression was conducted for each condition separately, such that average compositionality score predicted the proportion of compositional definitions that subjects selected for the definitional ($\beta = .693$, $t(42) = 6.22$, $p < .0001$) and the generation ($\beta = .41$, $t(31) = 2.51$, $p = .018$) conditions.

---

[2] A regression was also conducted to examine whether performance on the mapping task ($M = .58$) predicted the proportion of compositional definitions that subjects in the mapping condition selected. However, no relationship between these two variables was found, $t(34)\ 1.00$, $p = .322$.
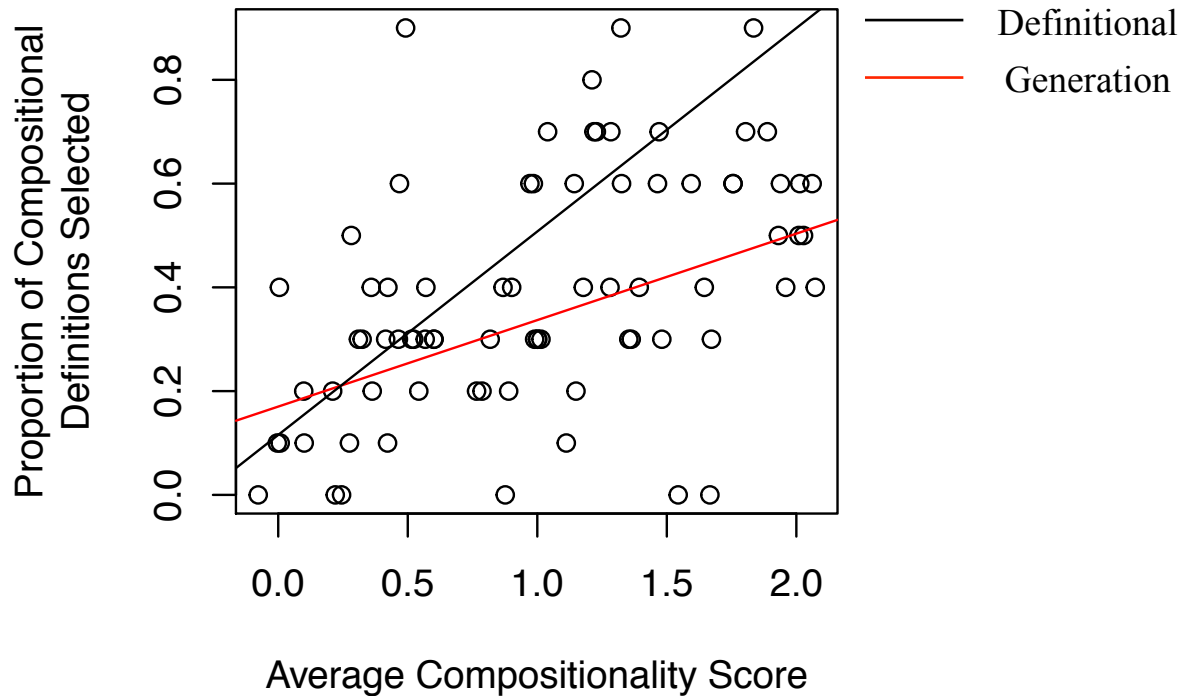
Figure 6. Illustrates the relationship between the proportion of compositional definitions that were selected and subjects' average compositionality scores, with the line of best fit for each condition in Experiment 1.

An additional regression revealed that there was an interaction between the definitional and generation conditions and subjects' compositionality scores, $\beta = .562$, $t(73) = 2.45$, $p = .017$, such that subjects' average compositionality scores were more predictive of the type of definition that subjects selected in the definitional condition than in the generation condition. One possible reason for this finding is that in the generation condition, subjects were required to use more of the nouns' component parts to create a scenario about each noun, whereas this was not necessary in the definitional condition. To further test this idea, two additional analyses (one for the definitional condition and one for the generation condition) were conducted that compared whether subjects' average compositionality scores differed from the average compositionality score (1.0) that would be expected if subjects' did not show a preference towards compositional-

or unitary-based responses. The analyses showed that subjects in the definitional condition were well under the expected value, ($M = .65$), $t(43) = 5.16$, $p < .0001$, whereas the opposite was true for subjects in the generation condition ($M = 1.47$), $t(32) = 3.39$, $p < .0001$. Furthermore, subjects in the generation condition ($M = 1.47$) had higher compositionality scores than subjects in the definitional condition ($M = .65$), $t(75) = 7.68$, $p < .0001$. Nevertheless, these differences did not produce differences between these two conditions in the type of definitions that were selected on the representation questions. Thus, responses from the definitional condition might be more reflective of how subjects typically represent the relational nouns that were used in this study.

One alternative to this possibility is that the definitional task is much closer to the representation questions than the generation task, and thus it is not surprising that responses on the definitional task were highly correlated with responses on the representation questions. However, these responses should only be correlated if the definitions that the subject was asked to select from were indeed correlated with the definitions that they wrote out. To the extent that the definitions that subjects wrote out provide a somewhat accurate assessment of how subjects represent the corresponding concepts, which appears to be a reasonable assumption, the strong relationship between the generated responses on the definitional task and the representation questions reaffirm the construct validity of this study's dependent measure. Hence, responses on the representation questions might provide a somewhat reasonable assessment of how subjects represent the corresponding concepts.

Although the findings that there are individual differences in how relational nouns are represented can be taken as support for the idea that relational concepts can indeed be represented in two fundamentally different ways (compositionally and unitarily), these findings are not conclusive. One potential issue that arises is that the compositional definitions were

longer in word length than definitions that were defined unitarily. Similarly, subject responses with higher compositionality scores tended to be longer in word length than those that had lower compositionality scores (i.e., responses that were more unitary-based). Thus, one alternative explanation for the present findings is that subjects who provided responses that received higher compositionality scores simply have a preference for descriptions that are wordier, and were thus more inclined to select wordier definitions. To address this possibility, two separate regressions were conducted, one for the definitional condition and one for the generation condition, which used the compositionality scores of these subjects to predict the type of definitions that they selected, while controlling for word length in these subjects' responses. These analyses showed that subjects' average compositionality scores in the definitional ($\beta = .70$, $t(41) = 3.31$, $p = .002$) and generation ($\beta = .51$, $t(30) = 2.03$, $p = .05$) conditions still predicted the type of definition that these subjects selected (i.e., subjects with higher compositionality scores selected a higher proportion of compositional definitions). These results lend further support to the idea that individual differences in the type of definitions that were selected were driven by corresponding differences in how subjects represent relational nouns, such that subjects selected the definitions that best corresponded with their representations.

Moving on, the lack of relationship between the Need for Cognition Assessment and the type of definitions that subjects selected suggests that the amount of enjoyment that subjects derive from engaging in reflective thought is not related to how they represent relational nouns. However, given the extensive familiarity that subjects had with the stimuli, it is unclear whether a different finding might be obtained with more novel relational concepts (e.g., mathematical equations, *Ohm's law*). One possibility is that people with a high need for cognition prefer unitary-based representations for relational concepts that are highly familiar (perhaps because the

concept has been overlearned), but compositional-based representations for more novel relational concepts which have yet to be fully learned.

Lastly, one possible explanation for the negative relationship between subjects' digit span score and the proportion of compositional definitions that were selected is that subjects with a higher verbal working memory (as measured through the forward digit span task) are better able to recognize that relationally structured information can be chunked into unitized concepts, which can be represented more efficiently than concepts that are represented compositionally. As a result, these subjects might develop a preference for unitary-based representations. Thus, differences in working memory capacity might give rise to differences in how relational concepts are represented. The inverse of this idea is also possible, such that subjects who are better at chunking consequently select unitary-based definitions and perform better on verbal working memory tasks. An alternative possibility that is perhaps more intriguing is that performance on the forward digit span task is facilitated by how subjects represent relational information, such that subjects who tend to represent relational concepts unitarily have more cognitive resources available to expend on a working memory task (than subjects who represent such concepts compositionally), which can thus lead to differences in performance among subjects on such tasks. In these latter two cases, differences in how subjects represent relational concepts gives rise to individual differences in working memory.

Traditionally, research on working memory has emphasized processing capacity as a means to explain individual differences in working memory (Melby-Lervåg & Hulme, 2013). However, the present findings raise the possibility that how relational information is represented can give rise to differences in how subjects perform on working memory tasks, such that subjects who use more efficient representations (e.g., unitary-based) perform better on these tasks than

subjects who represent relational concepts compositionally, as such representations might free up cognitive resources that facilitate performance on working memory assessments. This idea points to a topic that has not been given sufficient consideration in the research and theory on working memory, which is the role that representation plays in information processing and how such representations might affect subjects' performance on a working memory task.

Three somewhat antithetical explanations are put forth above to account for the finding that subjects with higher working memory selected a lower proportion of compositional-based definitions. Critically, all three hypotheses seem equally viable, as they can each account for the present findings equally well. More work will therefore be required to better test each of these accounts. Such work is of critical importance to the literature on individual differences in working memory, as it would help cognitive scientists to better understand the factors that contribute to these differences.

Experiment 1 Conclusion

The findings from Experiment 1 showed that subjects had a strong preference for definitions that were described unitarily. These findings also point to individual differences in how relational nouns are represented, such that some subjects seem to represent these concepts unitarily, whereas others represent them compositionally. One possibility is that these findings are simply a byproduct of the particular definitions that were used, such that the unitary definitions were simply better worded or were more accessible to subjects. Furthermore, it is possible that subjects simply preferred shorter definitions. However, such possibilities do not account for the fact that as subjects' compositionality scores (in the both the definitional and generation conditions) increased so too did the proportion of compositional definitions that they selected (as can be seen in Figure 6). Specifically, if subjects simply preferred the wording of the

unitary definitions or definitions that were shorter, there should not have been a positive relationship between subjects' compositionality scores and the proportion of compositional definitions that subjects selected. That is, unless subjects' generated responses and their responses to the representational questions were driven by the same underlying preference. It is however unclear exactly what that preference is (e.g., a representational preference or a preference for shorter descriptions).

Related to this latter possibility, subjects might have simply been better at writing out shorter responses (for definitions and scenarios) that consisted of synonyms than they were of writing out structured responses, and these subjects simply selected the definitions that most closely resembled their own generated responses. In such a case, it seems reasonable to argue that the generated responses actually reflect (to some degree) how subjects represent the corresponding concepts, such that the reason it was easier for subjects to generate unitary-based responses is because they actually represent the corresponding concepts unitarily. Moreover, one might expect that how a subject represents a given concept should be reflected (at least partially) in that subject's description of the concept, in much the same way as it is reasonable to expect that a student's knowledge about a given concept is at least somewhat reflected in their response to a question about that concept. Nevertheless, it is acknowledged that this argument is speculative and there are other possibilities as to why subjects might have generated the type of responses that they did. This uncertainty is a limitation of this study, but one that holds for all forms of assessment, as there are many possible reasons as to why a subject or student might respond in the way that they do and the actual reason cannot be known for certain.

This limitation not withstanding, Experiment 1 provides evidence for the idea that relational nouns can indeed be represented in two fundamentally different ways. If these findings

extend to most other relational concepts, they have the potential to have far ranging implications for various influential theories and models of relational learning (e.g., structure-mapping theory; Doumas, Hummel, & Sandhofer, 2008; Gentner, 1983; Hummel & Holyoak, 1997, 2003). For instance, as discussed above, structure-mapping theory holds that analogical learning and reasoning are driven by a process in which the corresponding elements between two analogous concepts are mapped and put into alignment in a way that preserves the elements' parallel connectivity, leading to the abstraction of the concepts' shared structure. However, if people typically represent such concepts unitarily, then there is no need for alignment, as two analogous concepts can be recognized as members of the same relational category if they share a category-defining concept that is unitized (e.g., both concepts are defined by a "*faster-than*" relation or by the attribute of *fast*; Corral et al., 2017; Glucksberg, 2003).

Although it is important to note that the present findings are far from conclusive, they nevertheless provide a strong reason to question the assumption that people typically represent relational concepts compositionally. This is a critical assumption on which many influential theories (e.g., structure-mapping theory, Gentner, 1983) and computational models of relational learning have been premised upon (e.g., DORA (Discovery of Relations by Analogy), Doumas et al., 2008; SME (Structure Mapping Engine), Falkenhainer et al., 1989; MAC/FAC (many are called, few are chosen), Forbus et al., 1995; LISA (Learning with Inference and Schemas and Analogy), Hummel & Holyoak, 1997, 2003; AMBR (Associative Memory-Based Reasoning), Kokinov, 1990, 1994). However, this assumption has not been rigorously examined empirically. The remainder of this paper focuses on addressing this issue and tests this assumption across six experiments, which use a variety of relational learning tasks and paradigms.

CHAPTER III

Experiment 2

Due to the representational flexibility that humans possess (Chalmers et al., 1992; French, 1997; Mitchell & Hofstadter, 1990), it seems plausible that relational concepts can be represented both unitarily and compositionally. Indeed, this idea was supported by the findings from Experiment 1. For instance, a subject might represent a concept such as *investigation* based on a global attribute (e.g., an inspection), but can also likely represent its relational substructure when necessary (explicitly representing the agent, question, line of inquiry, and their interrelations). This idea leads to the question of which type of representation people use by default when learning a relational concept. Although Experiment 1 suggests that subjects default to unitary-based representations, subjects were likely highly familiar with these concepts, and thus a different outcome might be observed in cases where subjects are required to learn novel relational concepts.

The main hypothesis of the present study is that, because unitary representations should allow for more efficient processing (because they are posited to be psychologically similar to features), subjects will use such representations when they are available. We test this prediction by giving subjects relational category learning tasks and encouraging them to represent the stimuli either compositionally or unitarily. If people typically learn relational concepts from structural alignment (Gentner, 1983; Gentner & Markman, 1997; Markman & Gentner, 2000), then encouraging subjects to use compositional representations should aid learning. However, if people instead learn more efficiently with unitary representations, than the opposite outcome should be expected.

Half the subjects in this study were given a classification task, in which they were shown a series of stimuli and were asked to make categorization judgments. Unitary representations seem especially well-suited for such a task, because they should enable subjects to directly recognize the diagnostic property in a stimulus, just as with feature-based categories. The other subjects were given an inference task, in which they were asked on each trial to determine a missing property of a stimulus that was presented together with its category label. Research with feature-based categories has shown that classification and inference learning tend to yield different category representations, with inference tasks encouraging learning of internal category structure, such as correlations among features (Markman & Ross, 2003; Yamauchi & Markman, 2000). Thus, inference tasks seem to strongly encourage compositional-based learning, whereas this is not necessarily the case for classification tasks. This finding suggests that compositional representations should be particularly well-suited for inference learning with relational categories, as such representations highlight the internal structure of stimuli. The inference conditions of our experiments thus provide a more stringent test of the hypothesis that people can learn relational concepts better through unitary representations.

This study hence extends the finding from Experiment 1 that people represent relational concepts in two fundamentally different ways (compositionally and unitarily) and investigates how these two types of representations can be leveraged to improve relational learning. More specifically, Experiment 2 examines how providing unitary and compositional descriptions (manipulated through the use of hints) of relational concepts affects learning on classification and inference tasks (description and task type both manipulated between subjects). Subjects were provided either a unitary or compositional hint at the start of learning and again after every third error, in order to assess whether each type of hint can improve learning. Control groups who

were given no hints were also included in order to assess baseline performance in both tasks and test whether each type of hint improves relational learning.

The stimuli used in this study were taken from Corral et al. (2017), which were adopted and modeled after those used by Rehder and Ross (2001). A stimulus consisted of three sentences, each of which describes a different component of a machine that works to remove waste material: (1) the location of where the machine operates, (2) the waste material the machine removes, and (3) the instrument the machine uses.

Stimuli were sampled from two categories: coherent and incoherent. Each category consisted of 18 exemplars. The categories were determined by how a machine's components were related to one another. For exemplars from the coherent category, the machine's instrument is suited for collecting the waste material that the machine works to remove, which can be found in the location where the machine operates. Consider the following example: "Operates on the seafloor, works to remove lost fishing nets, and has a hook." This exemplar is coherent because of the secondary relations among the machine's component parts (presumed to be known by subjects), such that lost fishing nets can be found on the seafloor and a hook can be used to retrieve lost fishing nets. In contrast, exemplars from the incoherent category do not satisfy these second-order relations (i.e., the machine's tool cannot be used to collect the machine's target waste material and that material cannot be found where the machine operates). Figure 7 illustrates the abstract relational structure of the two categories.

Half of the subjects completed an A/¬A classification task (in which each stimulus was to be categorized as either a category member or a nonmember), and the other half completed an inference task. On each trial, the subject was presented a single stimulus and was asked to make

an inference or classification judgment (depending on the condition). After the response, the subject was shown whether the response was correct along with the correct answer.

Operates on the surface of the water
Works to clean spilled oil
Has a spongy material

Operates on the surface of the water
Works to collect dangerous gaseous ions
Has a shovel

Coherent Item
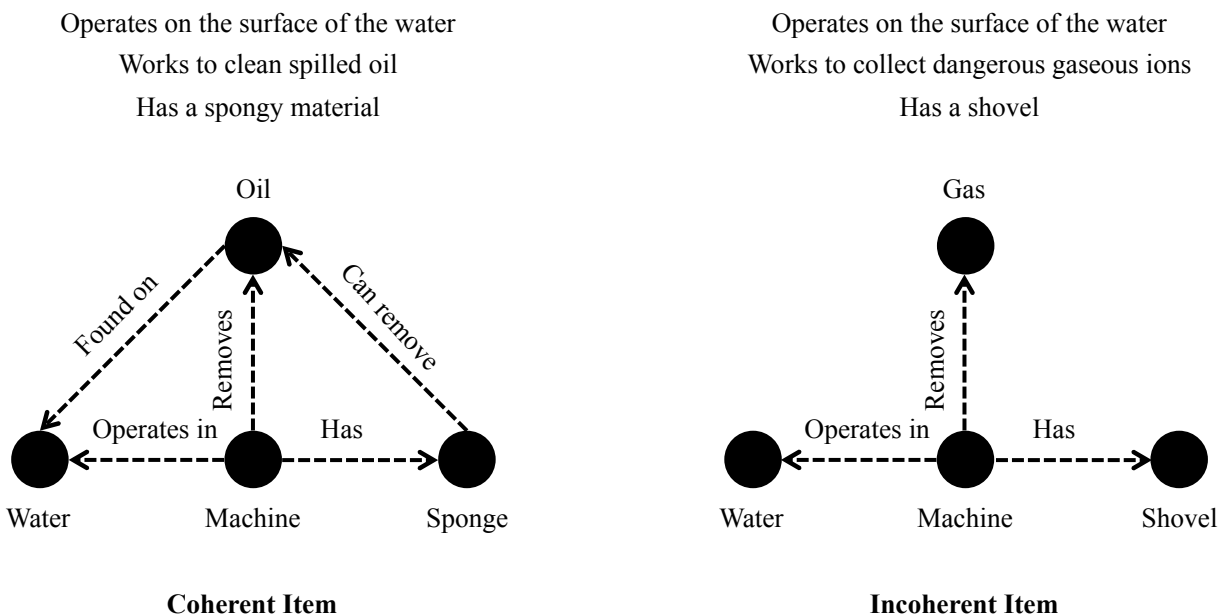
Incoherent Item

Figure 7. Illustration of the relational structure for items in the coherent and incoherent categories in Experiment 2. The structures differ in that coherent items satisfy the relations indicated by diagonal lines: the machine's implement can remove the target, and the target is found in the machine's location. Recreated from Corral et al. (2017).

Method

Participants

Two hundred eighteen undergraduates from the University of Colorado Boulder participated for course credit in an introductory psychology course. Subjects were randomly assigned to six conditions (between-subjects). Type of hint (compositional vs. unitary vs. control) was crossed with task type (classification vs. inference): compositional-classification (*N*

= 36), compositional-inference ($N = 32$), unitary-classification ($N = 37$), unitary-inference ($N = 37$), control-classification ($N = 37$), and control-inference ($N = 39$).

Stimuli

The stimuli for this study are included in Appendix B. Half of the stimuli for this study were taken from Rehder and Ross (2001) and Higgins (2012); the other half were taken from Corral et al. (2017). Rehder and Ross created three coherent items and three incoherent items; the incoherent items were generated by re-arraigning the features of the coherent items, in a way that each incoherent item took one feature from each of the three coherent items. Higgins used a similar method to generate an additional 12 items (six coherent and six incoherent). Eighteen additional items (nine coherent and nine incoherent) were created by Corral et al. (2017) by using the same method as Rehder and Ross. This study therefore consisted of 36 total stimuli, 18 from each category. Subjects were presented a single stimulus, which was shown as three lines of text surrounded by a red border, as shown in Figure 8.
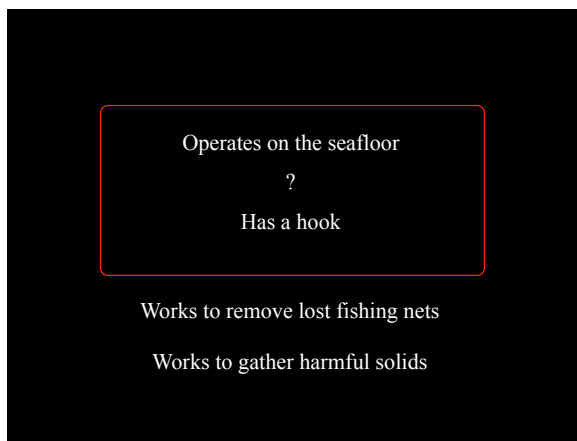
Figure 8. Example of a stimulus display from the coherent category (Morkels) from the inference task in Experiment 2.

Design and Procedure

All stimuli were presented on a 16-inch LCD monitor on a black background and all responses were entered using a computer keyboard. Subjects were told that they would be shown short descriptions of various types of cleaning machines, some of which were made by the Morkel Company (coherent category) and some were not (incoherent category). Additionally, subjects were told that the Morkel company makes many different types of cleaning machines, which operate in different environments, work to remove different types of materials, and use different types of tools. Subjects were provided a positive example of a Morkel (randomly selected) and were told that all Morkels share a certain commonality and it was their job to figure out what it was.

Subjects in the unitary condition were shown the following hint: "On each trial try to think about how "well suited" the machine is for performing its task. Keep in mind that consumers say machines from Morkels are built "intuitively" in a way that makes sense." This hint was intended to shift subjects' attention toward finding a global attribute of the stimulus and away from the explicit relationships among its components. Using this hint, it is possible for subjects to learn how to distinguish the categories without explicit knowledge of their relational structure. This hint can therefore be said to encourage subjects to represent the concepts unitarily.

In contrast, subjects in the compositional condition where shown the following hint: "On each trial try to think about the specific manner in which the machine's 1st property relates to its 2nd and 3rd properties, as well as how its 2nd property relates to its 3rd property." This hint was intended to focus subjects' attention on the relationships among the component parts of the stimulus, and thus to encourage them to represent the stimulus compositionally.

Subjects were presented the appropriate hint during the initial task instructions, after the first trial, during rest breaks, and following every third error that the subject committed (on a blank screen after corrective feedback was shown); the corresponding hint was presented every third error to strengthen the manipulation. Following every third error, the screen was cleared and the corresponding hint was presented in red bolded letters at the center of the screen. Subjects were asked to read the hint carefully and press the spacebar when they were ready to continue. Subjects in the control group were not shown a hint and were instead asked to continue to try their best; this reminder was presented during the initial task instructions, after the first trial, on every third error that the subject committed, and on rest breaks (as in the other conditions).

Each subject completed 72 trials. The order in which the items were presented was randomized for all subjects. After each block of 18 trials, subjects were given a self-paced rest break and were shown the proportion of correct responses they answered correctly over those trials, along with the number of trials they had completed and the number that remained. Subjects were also shown the corresponding condition hint. On each rest break, subjects were asked to press the spacebar when they were ready to continue.
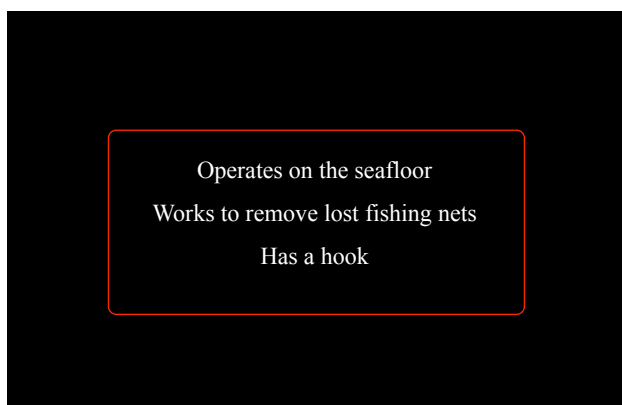
Figure 9. Example of a stimulus display from the coherent category (Morkels) from the classification task in Experiment 2.

Figure 9 shows a stimulus from the classification task. On each trial in the classification condition, a single, complete stimulus was presented and the subject was shown a prompt (directly above the stimulus at the center of the screen) that told them to type "A" if the machine was a Morkel or "L" if it was not. On each trial in the inference condition, the category label for a stimulus was shown (Morkel or non-Morkel) directly above an incomplete stimulus consisting of two of its three components (i.e., sentences). Below the stimulus were two response options, one of which was the missing component and the other was a lure. The component that the subject was asked to infer (i.e., implement, target material, or location) was randomly selected on each trial. Subjects were asked to select which was the missing component by typing "A" if the correct choice was the top option or "L" if it was the bottom option. The spatial position in which the two options were presented was randomized on every trial. For items that were Morkels, the correct response was the option that shared secondary relations with the given stimulus' components. The lure did not share secondary relations with either of the stimulus' components. This choice was made to maximally differentiate Morkels from non-Morkels. Thus, for items that were non-Morkels, the correct response was the component that did not share any secondary relations with either of the stimulus components. The accompanying lure shared at least one secondary relation with one of the stimulus' components. Figure 8 shows an example stimulus from the inference condition. For each stimulus, two inference lures were selected from the stimuli (beforehand) for each of the stimulus' components. On each inference trial, the corresponding inference lure was randomly selected from these two options.

After each response, subjects were provided corrective feedback, in which they were shown whether they were correct, along with the correct response. This feedback was presented at the center of the screen directly under the stimulus (which remained on the screen). Responses that were incorrect were shown in red and responses that were correct were shown in green. Feedback remained on the screen for 3 s and the intertrial interval was 400 ms.

At the end of the study, subjects were presented two options, each of which described the category rule for Morkels and non-Morkels. The category rules for one of the options were written unitarily, whereas for the other option the category rules were written compositionally. Subjects were asked to select the option that best represented how they were thinking about the two categories by typing "a" or "b". For each subject, the letter option for each type of description was randomly assigned. The unitary option read as follows: "Machines that were Morkels made sense or seemed like they would function, whereas machines that were non-Morkels did not make sense or seemed like they would not function". The compositional option read as follows: "For machines that were Morkels, what the machine worked to remove could be found where the machine operated, and could be removed with the tool the machine used, whereas for non-Morkels, what the machine worked to remove could not be found where the machine operated and could not be removed with the tool the machine used". After subjects entered a response, the screen was cleared and they were provided a thank you prompt at the center of the screen; the prompt remained on the screen for 500 ms. The study ran for an average length of approximately 15 min.

Results and Discussion

Figure 10 shows average learning curves for subjects in each group. An ANOVA was conducted to examine differences in performance among groups.[3] The analysis showed a main effect of hint, $F(2, 212) = 42.14$, $p < .0001$, $MSE = .014$, and an interaction, $F(1, 212) = 8.90$, $p = .0002$, $MSE = .014$, indicating that the main effect of hint depends on the type of task that subjects completed. On the classification task, control subjects ($M = .61$, $SE = .024$) were outperformed by subjects in the compositional ($M = .775$, $SE = .024$; $t(71) = 4.89$, $p < .0001$, $d = 1.60$) and unitary groups ($M = .830$, $SE = .016$; $t(72) = 7.75$, $p < .0001$, $d = 1.83$). In the inference condition, only subjects who received a unitary hint ($M = .716$, $SE = .014$) performed better than control subjects ($M = .585$, $SE = .016$; $t(74) = 5.34$, $p < .0001$, $d = 1.24$), as no differences were observed between subjects who were presented a compositional hint ($M = .587$, $SE = .018$) and subjects in the control group.

---

[3] To test for block by condition interactions, the data were also analyzed using a mixed-model ANOVA, with block as a within-subjects factor (each block consisted of 9 trials, totaling 8 blocks) and type of hint and task type as between-subject factors. However, no statistically reliable interactions were found between block and these between-subject factors, indicating that the results reported above did not depend on block. This model was also run for Experiments 3-7, but no statistically reliable interactions were found between block and any of the between-subject factors that were used in those studies. For this reason, this model is not discussed further.
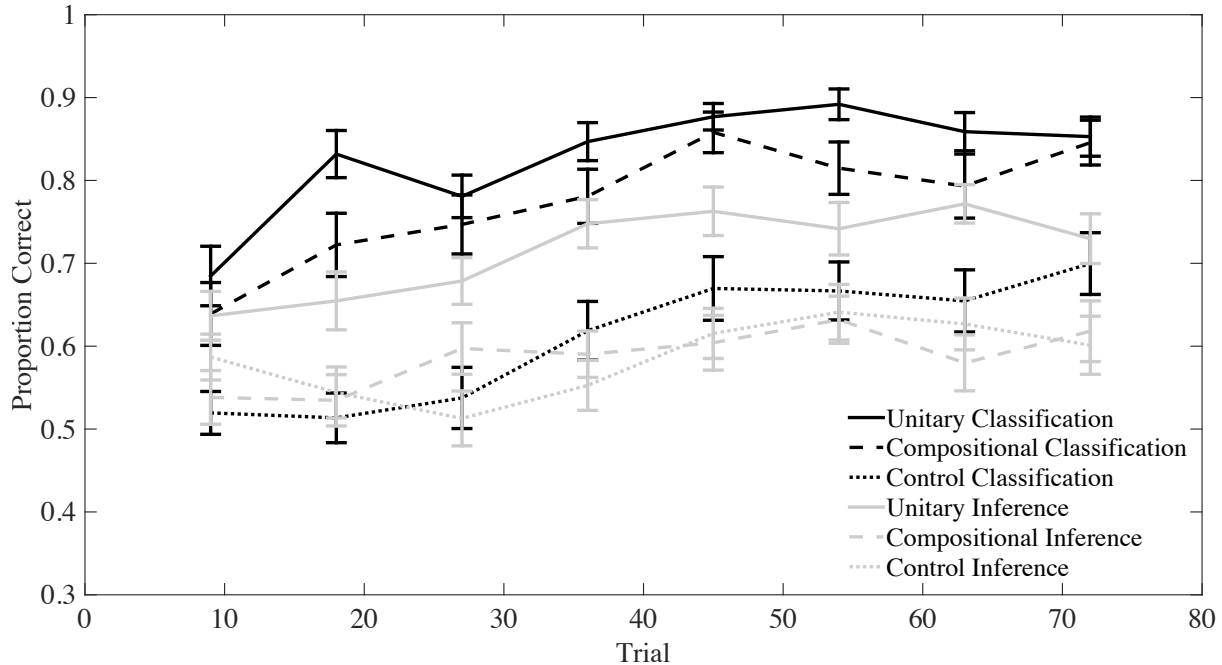
Figure 10. Average learning curves and standard errors across blocks of nine trials for each condition in Experiment 2.

Planned *t*-tests were conducted to compare the unitary and compositional groups, separately for each task. On the classification task, subjects in the unitary condition outperformed subjects in the compositional condition (*M* = .775, *SE* = .024), *t*(71) = 1.85, *p* = .068, *d* = .45. This same pattern was observed in the inference condition (unitary *M* = .716, *SE* = .014; compositional *M* = .587, *SE* = .018), *t*(67) = 5.28, *p* < .0001, *d* = 1.29. An additional 2 (unitary vs. compositional) × 2 (classification vs. inference) ANOVA was conducted, which excluded control subjects. This analysis revealed an interaction, *F*(1, 138) = 4.01, *p* = .047, *MSE* = .013, indicating that the unitary advantage was stronger in the inference task than in the classification task.

One possibility that is explored further here is that the learning advantage produced by the unitary hint is specific to the concept that the hint corresponds to and does not generalize to

aid subjects in discovering related concepts. In the present study, subjects were given a hint about the rule that defined the Morkel category, but were not provided explicit information about non-Morkels. Thus, it is possible that subjects who received a unitary hint only learned the category rule for Morkels, but not for non-Morkels. Indeed, subjects could perform exceptionally well on the classification task by simply learning the category rule for Morkels, without having explicit knowledge of the non-Morkel category rule, as non-Morkels could be identified by the absence of a Morkel. In contrast, this strategy is not as useful in the inference task, as subjects must infer the missing component for Morkels and non-Morkels alike. This idea can thus be tested by comparing subjects' inference performance on trials that consisted of non-Morkels.

Figure 11 shows mean overall performance across all trials on the inference task for Morkel and non-Morkel trials. First, a mixed-model ANOVA was conducted, with trial type as a within-subjects factor and type of hint as a between-subjects factor. A main effect of hint was found, as reported above. Furthermore, consistent with findings from Rehder and Ross (2001), there was a within-subjects main effect, $F(1, 217) = 115.78$, $p < .0001$, such that subjects performed better on items that were Morkels ($M = .75$) than items that were non-Morkels ($M = .619$). Critically, no interaction was found between type of hint and trial type, $p > .05$, indicating that the main effect of hint does not depend on trial type. Planned comparisons showed that subjects who received a unitary hint outperformed subjects who received a compositional hint on Morkel ($M_{unitary} = .811$; $M_{compositional} = .722$; $t(67) = 3.48$, $p = .0009$) and non-Morkel items ($M_{unitary} = .62$; $M_{compositional} = .451$; $t(67) = 4.58$, $p < .0001$). To be thorough, additional analysis were conducted, which showed that subjects who received a unitary hint also outperformed subjects in the control group on both Morkel ($M_{control} = .71$; $t(74) = 3.87$, $p = .0002$) and non-Morkel items ($M_{control} = .463$; $t(74) = 4.66$, $p < .0001$). No differences were found between

subjects who received a compositional hint and control subjects on either trial type, $ps > .61$. Taken together, these findings show that the unitary hint is useful for learning the concept that the hint pertains to and moreover, that such hints can be leveraged by subjects to discover novel relational concepts.
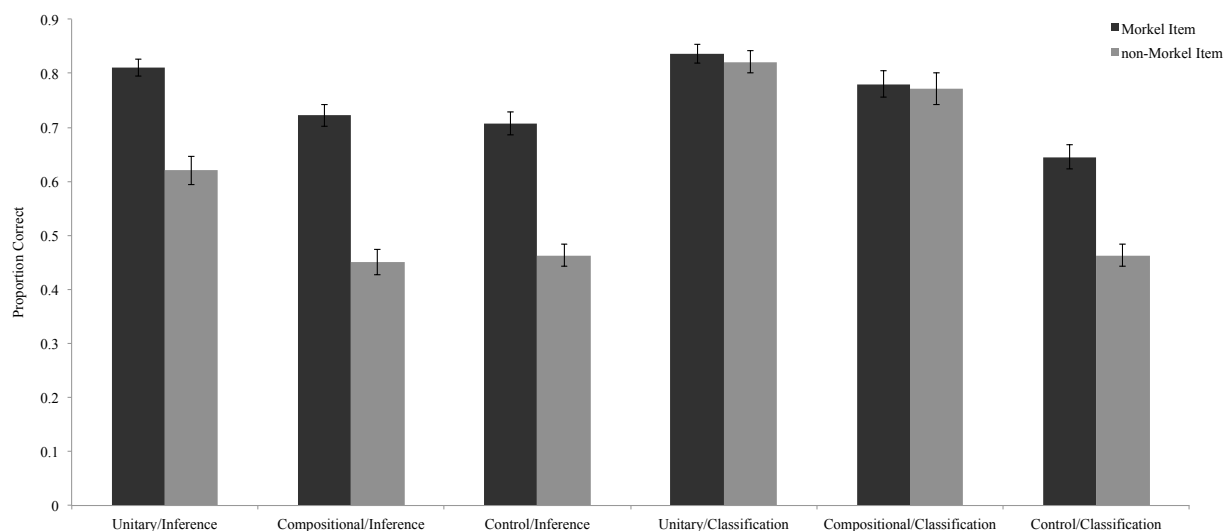
Figure 11. Mean performance in each condition on Morkel and non-Morkel items in Experiment 2 and the standard error of the mean.

Representational Preference

Figure 12 shows the proportion of observed responses in each condition for the compositional and unitary category rules. A binomial test was conducted to examine whether subjects had a preference between the two types of category rules (shown to subjects at the end of the study), which were either described unitarily or compositionally. Collapsing across conditions, a greater proportion of subjects selected the category rules that were described unitarily ($M = .61$) than category rules that were described compositionally ($M = .39$), $p = .001$. These findings are in line with the results from Experiment 1, and add further support to the idea that people might typically prefer to represent relationally structured concepts compositionally.

Additionally, a 3 (type of hint) × 2 (task type) logistic regression was also run to examine whether the number of subjects that selected each category rule differed by condition. This analysis showed a main effect of task type, β = -.556, $\chi^2(1)$ = 3.94, p = .047, as a greater number of subjects in the inference condition selected the category rule that was described unitarily than subjects in the classification condition; there was not a statistically reliable effect of hint nor was there a statistically significant interaction. This result is consistent with the performance differences (i.e. a greater unitary advantage for the inference task). One explanation for this finding might be that during an inference task, subjects attempt to find a unitary-based representation that they can use to guide their inferences, and thus show a higher preference for the category rule that is described unitarily than subjects in the classification condition.
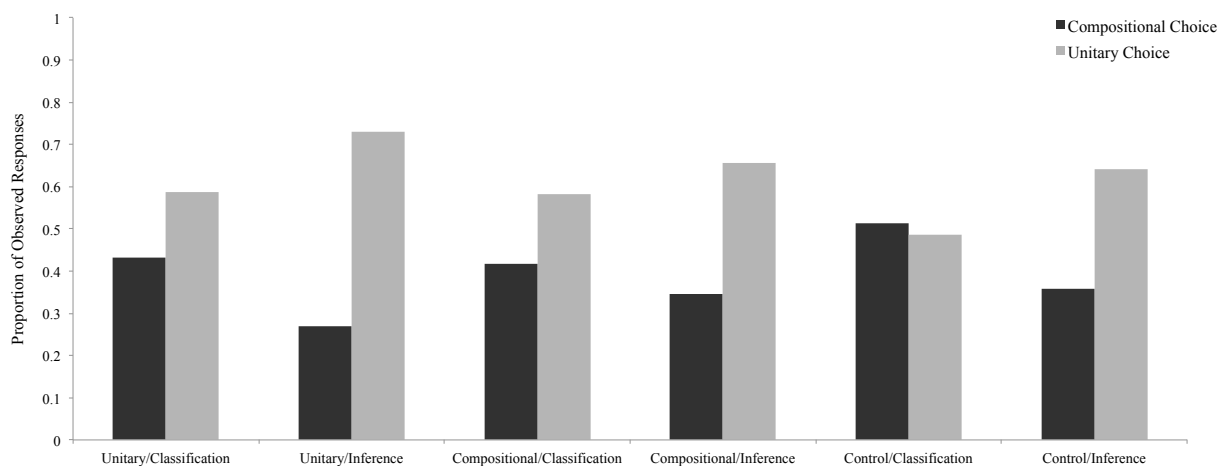
Figure 12. Proportion of observed responses of the compositional and unitary category rules in each condition in Experiment 2.

Individual Differences

Figure 13 shows individual learning curves for Experiment 2. Although these learning curves are fairly noisy, Figure 13 seems to show three primary types of learning: (1) partial

learning, (2) all-or-none learning, and (3) gradual learning. These different patterns in learning likely reflect individual differences among subjects, which might range from differences in representational preferences to differences in subjects' motivation and intelligence. The individual learning curves in Experiments 3-7 are similar to those in Figure 13 and exemplify that the stimuli and manipulations used in these studies can give rise to various types of learning patterns (neither of which was particularly predominant in any of the studies), which differ among subjects.
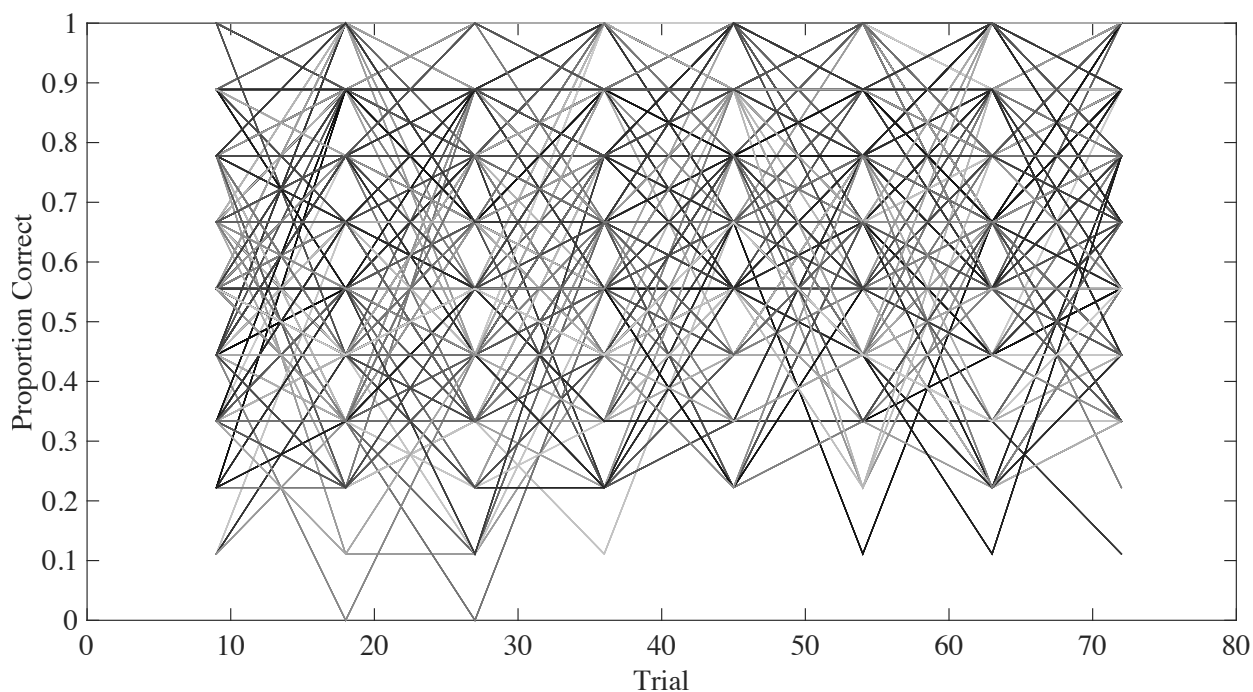
Figure 13. Individual learning curves for all subjects in Experiment 2, based on blocks of 9 trials.

Moving on, it is important to note that if there are individual differences in how subjects represent relational concepts, it is possible that the type of hint that is most effective varies by subject. Specifically, subjects who prefer to represent relational concepts compositionally might be more likely to gain a greater benefit from receiving a compositional hint than one that is

unitary, whereas the opposite might be true of subjects who prefer unitary-based representations. Importantly, the type of category rule that subjects selected at the end the Experiment 2 might provide insight into how they prefer to represent the two relational categories that were used in the study.

However, a regression showed that there was no relationship between the type of category rule that subjects selected and performance, $t(216) = .66$, $p = .512$. Additionally, a 2 (type of hint (excluding control subjects): unitary vs. compositional) × 2 (task type: classification vs. inference) × 2 (category rule: unitary vs. compositional) ANOVA was conducted. The results showed a non-significant trend in the interaction between type of hint and category representation, $F(1, 134) = 2.39$, $p = .12$, $MSE = .03$, suggesting that the type of hint that is most effective for learning might depend on how subjects typically represent a given relational concept. Specifically, for subjects who received a unitary hint, those who selected the unitary-based category rule ($M = .806$) marginally outperformed subjects (collapsing across task type) who selected the category rule that was described compositionally ($M = .753$), $t(72) = 1.91$, $p = .059$, whereas no differences were found between the type of category rule that was selected for subjects who received the compositional hint ($M_{\text{unitary rule}} = .68$; $M_{\text{compositional rule}} = .69$), $t(66) = .20$, $p = .84$.

A.

B.
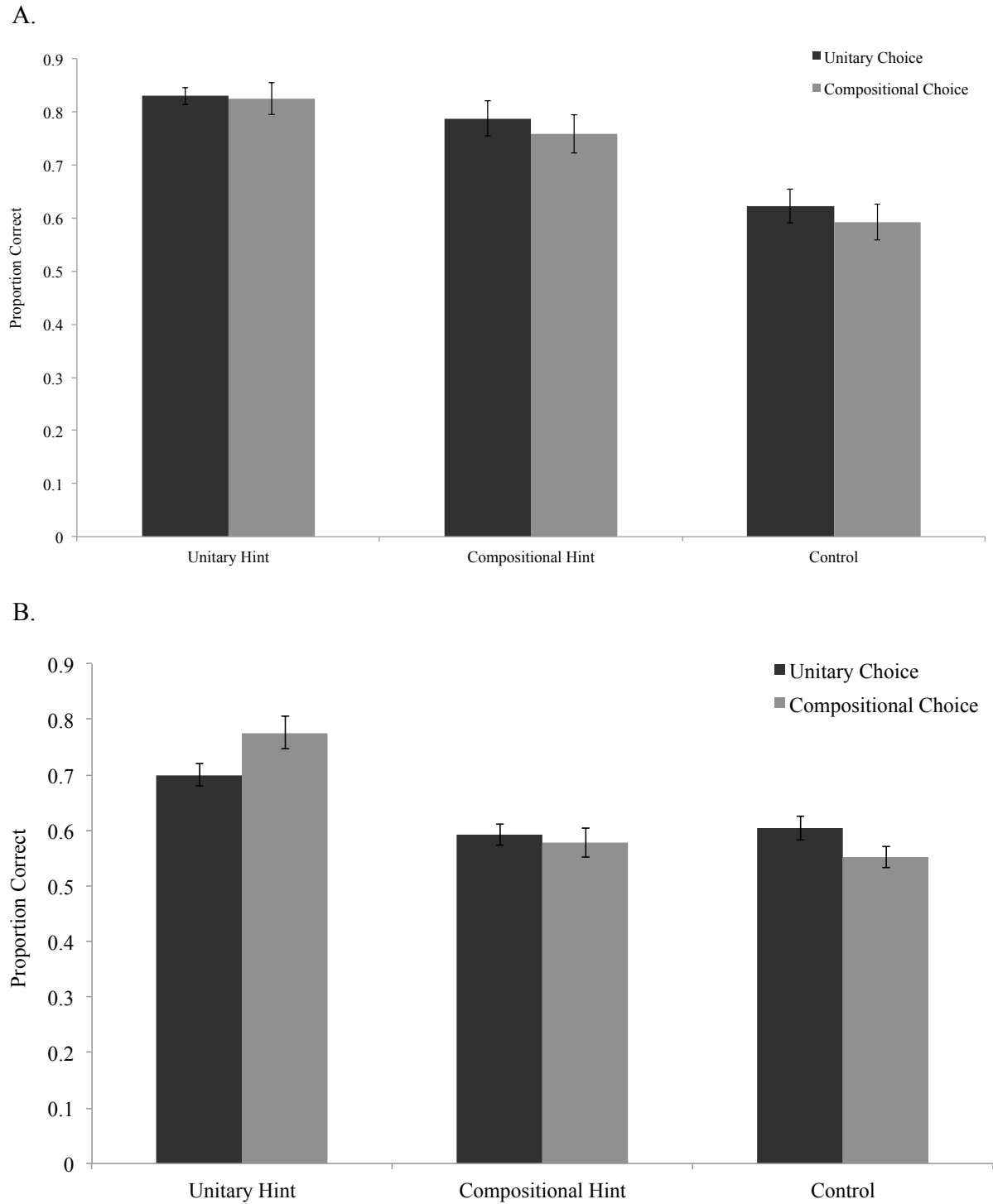
Figure 14. Mean performance of subjects in each group and task in Experiment 2, based on the type of category rule that was selected, along with the standard error of the mean. A: Mean performance on the classification task. B: Mean performance on the inference task.

To better understand this trend in the data, a 2 (type of category rule selected) × 2 (task type) ANOVA was conducted which was restricted to subjects who received a unitary hint. The analysis showed a marginal interaction between the type of category rule that subjects selected and task type, $F(1, 70) = 3.04$, $p = .086$, $MSE = .031$, such that for subjects who received a unitary hint, those who selected the compositional-based category rule ($M = .776$) outperformed those who selected the unitary-based category rule ($M = .693$) on the inference task, $t(35) = 2.11$, $p = .04$, but not on the classification task, $t(35) = .14$, $p = .89$. The opposite pattern of results was observed for subjects in the control group, as subjects who selected the category rule that was described unitarily marginally ($M = .60$) outperformed subjects who selected the compositional-based category rule ($M = .55$) on the inference task, $t(37) = 1.64$, $p = .109$, but no differences were observed on the classification task based on the type of category rule that subjects selected, $t(37) = .75$, $p = .461$; for subjects who were provided a compositional hint, no differences in performance were found based on the type of category rule that they selected on either task, all ps > .64. Figure 14 shows the mean performance of subjects in each group and task, based on the type of category rule that they selected.

It is important to note that in order to make the correct inference on the inference task, subjects were required to figure out the relationships among a stimulus' (i.e., machine's) component parts. Thus, one way to interpret the findings presented in this section is that for subjects who received a unitary hint, those who preferred to represent the stimuli compositionally were better able to use that hint to build a compositional representation of the stimuli than subjects who preferred to represent the stimuli unitarily. However, in cases where no hint was provided, subjects who prefer to represent the stimuli unitarily seem to perform marginally better than subjects who prefer to represent it compositionally. One reason for this

finding might be that in the absence of any type of hint, representing relational stimuli unitarily aids learning on an inference task. It therefore follows that the type of benefits that are conferred to subjects by representing relational concepts unitarily or compositionally might depend on whether subjects are shown a unitary hint.

It is important to remind the reader that subjects were asked to select between the category rules after completing the learning task and these responses therefore likely reflect how subjects learned the task. One possibility is that subjects' selection of the category rules was driven by how readily the category rules could be represented or that they selected such category rules because it was in line with how they typically represent relational information. A related possibility is that subjects learned the category rules in the way that they did because it was more accessible to them than the other category rules. In either of these cases, the same conclusion can be drawn, which is that subjects had a preference for one of the two category rules (either because it was in line with their representational preference or because it was more accessible), which is reflected by how they learned the task. Moreover, differences in performance that were based on the category rules that subjects selected might reflect which type of category rules (and representations) were more useful for learning the task. However, this latter conclusion is based on correlational data and is therefore far from conclusive and should be taken with caution.

Experiment 2 Conclusions

To return to the primary results, one surprising finding from this study was that the unitary advantage was stronger for the inference task than for classification. The effect size for the inference task was actually quite dramatic (Cohen's *d* of 1.29). It had been predicted that, if anything, the interaction would go in the opposite direction, given that inference tasks encourage learning the relationships among a concept's components (Markman & Ross, 2003; Yamauchi &

Markman, 2000). One speculative possibility is that inference learning encourages a top-down approach, in that subjects must reason from the category label to the stimulus, whereas classification encourages a bottom-up approach of reasoning from the stimulus to the category label. Likewise, a unitary representation is top-down in that it embodies a global property of a stimulus that can be used to deduce its internal structure, whereas a compositional representation is bottom-up in that the local structure is explicitly represented and the global property emerges only implicitly from the relational system. Under this view, there might be a congruency effect between the stimulus representation and the processes involved in carrying out the task. In particular, a unitary representation might be more congruent with an inference task, because it facilitates conceiving of a concept by a single attribute that can then be used to infer missing parts of a stimulus.

These speculations aside, the main conclusion of Experiments 2-3 is that although relational concepts are defined by the interconnections among their component parts, subjects seem to learn these concepts better when they can be represented unitarily, which might facilitate a global understanding that is easier to discover and use than an explicitly structured one. Furthermore, although compositional-based instruction can help subjects classify a given concept, it might not be optimal for inference-based reasoning.

CHAPTER IV

Experiment 3

Experiment 3 builds on the findings from Experiment 2 and examines how category learning is affected when subjects are encouraged to represent a relational concept one way (either unitarily or compositionally) and are subsequently made aware of an alternative representation. This manipulation examines how having access to both types of representations affects learning, which is an important question because people might often represent a given relational concept both unitarily and compositionally. The present study uses the stimuli from Experiment 2, and all subjects performed the classification task. All subjects were either provided a unitary or compositional hint prior to the start of learning. For half of the subjects, the hint was changed after the 18th trial (i.e., the unitary hint was replaced with the compositional one and vice versa); this change occurred midway through the full stimulus set. For the other half of subjects, the hint they were shown remained the same throughout the study. These latter conditions were identical to the unitary and compositional classification conditions in Experiment 1.

Method

Participants

One hundred fifty-seven subjects participated for course credit in an introductory psychology course. Subjects were randomly assigned to four conditions: unitary/switch ($N = 40$), compositional/switch ($N = 39$), unitary/no-switch ($N = 39$), and compositional/no-switch ($N = 39$). For each condition, the name of the hint indicates the type of hint that subjects were initially presented at the start of the study. Thus, subjects in the unitary/no-switch and the unitary/switch

conditions started the study with a unitary hint, whereas subjects in the other two conditions were initially shown a compositional hint.

Procedure

Following the same procedure as Experiment 2, subjects were presented a given hint (depending on their condition) at the start of the study, which was presented after the 1st trial and following every 3rd error (as in Experiment 2). After the 18th trial (i.e., following the first rest break), the screen was cleared and subjects in the switch conditions were shown the following prompt along with the other hint: "A new report from consumers indicates that a NEW method for distinguishing between machines that are Morkels and non-Morkels has been discovered". The hint was presented directly below this prompt at the center of the screen in bolded red font and subjects were asked to press the spacebar when they were ready to continue; subjects were also encouraged to try and use this new method of thinking about each machine on each trial. Following the 19th trial, the hint was presented once more and subjects were reminded to use it to try to figure out what constitutes a Morkel. Subjects in the switch conditions were shown this hint for the remainder of the study (i.e., on rest breaks and following every 3rd error), whereas no-switch subjects continued to see the hint they had seen at the beginning of the study. The average length of the study was approximately 15 minutes. The rest of the procedure was identical to that of Experiment 2.

Results and Discussion

Figure 15 shows average learning curves for subjects in each condition in Experiment 3. A *t*-test showed that subjects in the unitary/no-switch condition (*M* = .790, *SE* = .019) outperformed subjects in the compositional/no-switch condition (*M* = .716, *SE* = .028), *t*(76) =

2.23, p = .03, d = .50.[4] This finding directly replicates the results from the classification condition in Experiment 1, which showed a unitary learning advantage.
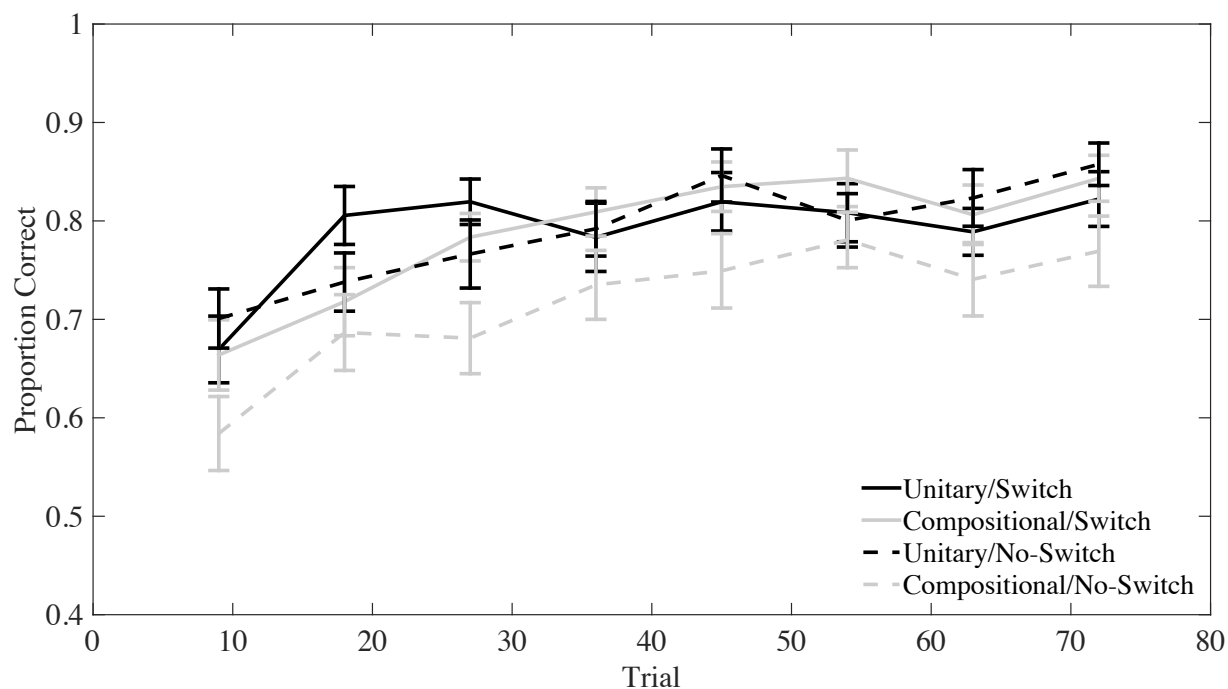
Figure 15. Average learning curves and standard errors of the mean across blocks of nine trials for each condition in Experiment 3.

In addition to this analysis, a series of planned comparisons were conducted to examine differences among groups from the point at which subjects were introduced to the other hint (trials 19-72). The first analysis showed that subjects in the compositional/switch condition (M = .82, SE = .018) outperformed subjects in the compositional/no-switch condition (M = .743, SE = .029), $t(76)$ = 2.26, p = .027, d = .51. Additionally, subjects in the unitary/switch condition (M =

---

[4] A model was run with phase (phase 1 (trials 1-18) vs. phase 2 (trials 19-72)) as a within-subjects factor and type of hint and type of switching as between-subject factors; phase did not interact with type of hint or type of switching.

.802, *SE* = .022) marginally outperformed subjects in the compositional/no-switch condition, *t*(77) = 1.77, *p* = .08, *d* = .45. However, no differences in performance were observed among any of the three groups that were presented a unitary hint at some point in the study. Thus, it seems that as long as a unitary hint is presented, regardless of whether it is the only hint that is shown or if it is presented before or after a compositional hint, subjects are able to benefit from it. Taken together, these findings support the conclusion from Experiment 1 and suggest that subjects indeed learn better when they rely on unitary representations.

Representational Preference and Individual Differences

As in Experiment 2, a binomial test (collapsing across all subjects) revealed that a greater proportion of subjects selected the category rules that were described unitarily (*M* = .68) than those that were described compositionally (*M* = .32), *p* < .0001. Figure 16 shows the proportion of observed responses in each condition in Experiment 3. These results further replicate the findings from Experiments 1 and 2 and suggest that people typically prefer to think of relational concepts in a unitary-based manner. Additionally, a logistic regression showed that there was a main effect of hint, β = -1.51, $\chi^2(1)$ = 6.85, *p* = .009, such that a higher proportion of subjects who started with a compositional hint (switch and no-switch condition) selected the category rules that were described unitarily. An interaction was also found, β = -.405, $\chi^2(1)$ = 4.15, *p* = .042, indicating that the effect of hint depends on the switch condition. Specifically, subjects selected the category rules that were described unitarily at a higher frequency in the compositional/switch condition than in the other three conditions (as shown in the Figure 16). It is important to note that subjects in the compositional/switch condition were first presented the compositional hint, and were then presented the unitary after the 18th trial, which was presented for the remainder of the experiment. This finding might therefore not be particularly surprising

given that these subjects were presented the unitary hint for most of the study. On the other hand, subjects in the other three conditions were shown either a unitary or compositional hint for as long or longer (true of subjects in the no-switch conditions) as subjects in the compositional/switch condition were shown the unitary hint. Nevertheless, subjects in these other three conditions did not show as strong of a preference for the category rules that were described unitarily as subjects in the compositional/switch condition. One explanation for this finding is that because subjects do not typically represent relationally structured concepts compositionally, it is more challenging for them to use these types of representations to learn a resulting concept. As a result, these subjects might have been hesitant to fully adopt a compositional representation. However, once these subjects were shown the unitary hint, which is perhaps a more natural representation that is easier to apply during relational learning, they fully adopt it and abandon the representation that was encouraged by the previous hint. The compositional/switch condition thus may have provided subjects the opportunity to experience the challenge of using a compositional representation during relational category learning, which was followed by the ease (and subsequent success) of applying the unitary hint.
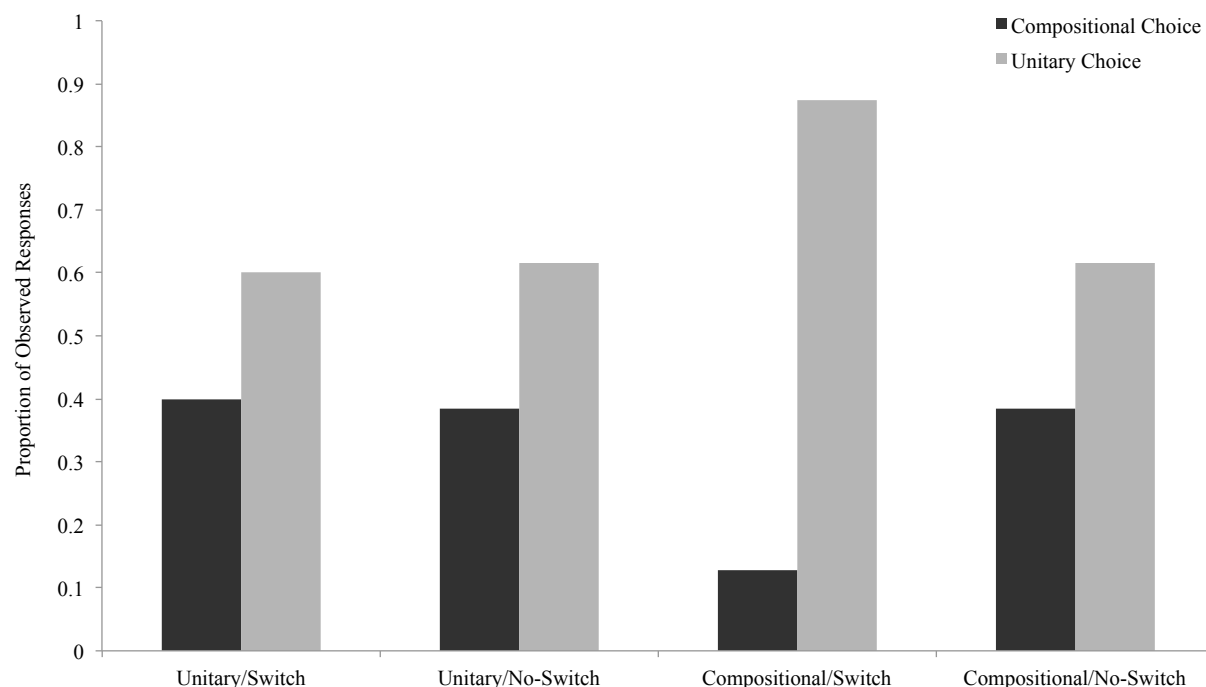
Figure 16. Proportion of observed responses of each category rule in each condition in Experiment 3.

This contrastive experience might have led these subjects to form a stronger preference for the unitary-based representation than subjects in the other three conditions, which might not have had this experience. Indeed, subjects in the compositional/no-switch condition were not exposed to the unitary hint and may therefore not have been fully aware of its benefits or the ease with which the stimuli could otherwise be represented. Subjects in unitary/no-switch condition were not presented a compositional hint, and as a result, these subjects might not have been aware of the difficulty of using this type of representation during category learning. Furthermore, although subjects in the unitary/switch condition were presented with both types of hints, these subjects were shown the compositional hint after the unitary hint, which might have helped to scaffold learning and buffer the difficulty of using the compositional hint, whereas subjects in

the compositional/switch condition did not have this buffer when they used the compositional hint.

Subjects within each condition were divided by the type of category rules that they selected and their performance was compared to one another (e.g., subjects' performance in the unitary/switch condition who selected the unitary category rules were compared to those in the unitary/switch condition who selected compositional category rules). However, no differences in performance were found among the groups. This finding is in line with the findings from Experiment 2, which found no differences in performance among groups on the classification task based on the type of category rules that subjects selected. One possibility is that performance on a relational classification task is not related to how subjects prefer to represent the corresponding relational concept. To further elaborate, it is possible that a classification task does not require subjects to build up and construct a representation in the same way that an inference task does (in which Experiment 1 found a relationship between the category rules that subjects selected and performance). Thus, the type of category rule that can be more readily represented might not be as relevant for classification as it is for inference. For this reason, performance on a classification task might not be affected by how subjects prefer to represent a relational concept in the same way as in an inference task.[5] An alternative explanation is that

---

[5] One finding that might be related to the type of task that was used in this study is that no differences were found in performance on Morkel and non-Morkel items (unlike in Experiment 2), $p = .93$. This finding is perhaps due in part to this study only using an A/¬A classification task, as subjects could presumably perform at a high level (on either item) by only learning the category rule for Morkels and classifying a stimulus as a non-Morkel when it failed to meet the appropriate criterion.

both compositional and unitary category rules can be equally effective in aiding subjects'

classification judgments, and thus how the category was learned did not affect performance.

Nevertheless, the findings presented in this section add further support to the idea that subjects

prefer (either due to representational or learning preferences) to represent relational concepts

unitarily.

CHAPTER V

Experiment 4

This study follows up on the work from Experiments 2-3 and examines how making both types of representations available (between-subjects) to subjects during relational learning might interact with classification and inference (within-subjects). Experiment 2 showed that although a compositional hint can be useful for category learning, it does not seem to aid inference. One explanation for these findings is that the inference task required subjects to engage in bottom-up learning, as they needed to discover the relationships among a stimulus' components in order to make the correct inference. Similarly, the compositional hint required the same type of bottom-up processing, and may therefore not have been particularly informative to subjects who completed the inference task. To elaborate, on each inference trial subjects were required to select between two options, one of which was the machine's missing component. Although the compositional hint directed subjects to look for the relationships among the machines' component parts, there are many potential relationships between each inference option for a given stimulus and its two component parts that were present in the stimulus. Thus, the compositional hint might not have provided the conceptual constraint that was necessary (and which might be provided by a unitary hint) to be useful for an inference task. In contrast, this issue did not arise on the classification task, as subjects were indeed capable of using the compositional hint to learn how to distinguish Morkels from non-Morkels. Accordingly, relational classification is more constrained than inference because a stimulus contains all of its properties and subjects must simply discover how those properties are related. One possibility is that a compositional hint can be used effectively during inference if a subject has a sufficient grasp of the concept that is being learned. One possible way to help subjects reach this goal is by

first training them on a classification task using a compositional hint and then asking them to complete an inference task. The present study tests this idea, along with how each type of hint affects inference learning when they are both made available to subjects.

This study used the same stimuli as Experiments 2 and 3, and a similar design and procedure to that of Experiment 2. First, all subjects performed a classification task, and were presented a given hint (either unitary or compositional, based on the condition) at the start of the study (as in Experiments 2 and 3). After responding to half of the stimuli, all subjects were given an inference-based task, at which point half of the subjects were presented a different hint and the other half were shown the same hint (as in Experiment 3). There are two primary benefits to this design. The first is that the classification task provides a measure of each hint's effectiveness. The second is that by presenting an inference task and then switching the hint that half of the subjects are shown, a direct measure can be obtained of how well subjects can transfer their relational category knowledge from classification to inference, along with which type of hint is more conducive to such transfer.

The primary prediction for this study is that subjects who are presented a compositional hint during classification and then switch to a unitary hint during inference (compositional/switch condition) should show the greatest improvement in performance when they shift from classification to inference. This prediction is premised on the idea that the compositional hint during classification should lead to lower performance than the unitary hint (as was shown in Experiments 2 and 3). However, once subjects in the compositional/switch condition are shown a unitary hint, the effectiveness of this hint should lead to a rapid improvement in learning. In contrast, subjects who are provided a unitary hint to start the study will likely develop a strong grasp of the category rules during classification and are therefore

more likely to display more stable performance when they switch to the inference condition than subjects in the compositional/switch condition. Another prediction that might follow from Experiments 2 and 3 is that subjects in the unitary/no-switch condition should perform better on the inference task than subjects in the compositional/no-switch condition. However, an alternative prediction is that the classification task will help subjects who receive a compositional hint better develop a sufficient understanding of the category rule, which will allow subjects in the compositional/no-switch condition to effectively transfer this knowledge to the inference task. As a result, no differences on inference performance might be observed between subjects in the compositional/no-switch condition and subjects in the unitary/no-switch condition.

Method

Participants

One hundred sixty-nine subjects participated in this study for course credit in an introductory psychology course at the University of Colorado Boulder. Subjects were randomly assigned to four conditions: unitary/switch ($N = 41$), compositional/switch ($N = 42$), unitary/no-switch ($N = 44$), and compositional/no-switch ($N = 42$).

Procedure

All subjects first completed 18 trials on a classification task. After the rest break on the 18th trial, the screen was cleared and subjects were shown a prompt that notified them that they would be completing an inference-based task (which they were given for the remainder of the study), and were told to press the spacebar to continue. This prompt was followed by a prompt that reminded subjects of the hint that they were already using or a new hint, which informed

them that there was a different way in which the concept could be represented. The rest of the design and procedure is identical to the inference task that was used in Experiment 2. The average length of the study was approximately 17 min.

Results and Discussion

Figure 17 shows learning curves for each condition on each task type. As in Experiments 2 and 3, subjects performed better on items that were Morkels (M = .767) than non-Morkels (M = .625), $t(168)$ = 15.09, p < .0001, SE = .001, d = 2.33, but no interaction was found between trial and task type or among trial type and the other two between-subject factors (i.e., type of starting hint and type of switching), p > .05; trial type was therefore not included in further analyses. A mixed ANOVA was run with task type as a within subjects-factor and type of starting hint and type of switching as between-subject factors. The results showed no between-subject main effects, but an interaction was found between task type and type of starting hint $F(1,165)$ = 4.30, p = .04, MSE = .011, such that subjects who were presented the compositional hint at the start of the study showed a greater increase in performance ($M_{increase}$ = .04) from the classification (trials 1-18) to the inference task (trials 19-72) than subjects who were shown the unitary hint ($M_{increase}$ = -.01) at the start of the study. Furthermore, a non-significant trend was observed such that subjects who were shown a unitary hint (M = .706) at the start of the study performed better on the classification task than subjects who were shown a compositional hint (M = .663), $t(167)$ = 1.60, p = .11, SE = .013, d = .248. However, these differences in performance were not present among conditions on the inference task, all ps > .15, suggesting that once subjects were introduced to a unitary hint, they were capable of making up any differences that might have developed in learning between them and subjects who were shown the unitary hint at the start of the study; no statistically reliable three-way interactions were found.
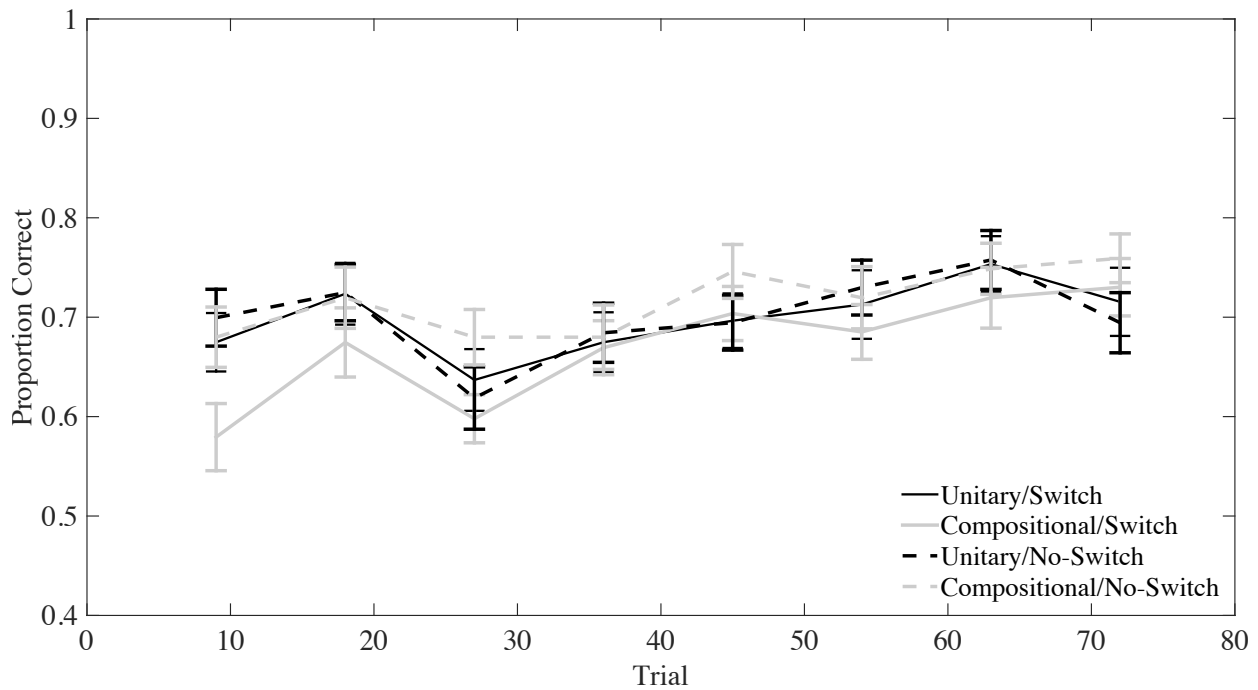
Figure 17. Average learning curves and standard errors of the mean across blocks of nine trials for each condition in Experiment 4. The first 18 trials indicate performance on the classification task; trials 19-72 indicate performance on the inference task.

Additional analyses revealed that subjects who were in the compositional/switch condition performed better on the inference task (M = .684) than on the classification task (M = .627), $F(1, 41) = 4.27$, $p = .045$, $MSE = .016$. Critically, this improvement in performance occurred after the hint was switched from compositional to unitary. Moreover, no statistically reliable improvements in performance were observed from the classification to the inference task for subjects in the other conditions, all $ps > .32$. The improvement in performance from the classification to the inference task was also greater for the compositional/switch condition than for the subjects in unitary/no-switch condition, $t(84) = 2.18$, $p = .032$, $SE = .017$, $d = .48$, as well as subjects in the unitary/switch condition (non-significant trend), $t(84) = 1.62$, $p = .109$, $SE =$

.018; a non-significant trend was also observed between the compositional/switch and the compositional/no-switch conditions, $t(82) = .98$, $p = .33$, $SE = .018$.[6] Taken together, these findings suggest that switching from a compositional to a unitary hint is more beneficial for relational learning.

However, it is important to note that in Experiment 2 there was a strong benefit to using the unitary hint (over one that was compositional) on the inference task. However, no such benefits were observed in the present study. One explanation for these somewhat incongruent results is that using compositional representations for inference-based learning (without classification training beforehand) can be very challenging, as subjects must build up a representation out of a relationally structured stimulus that is incomplete. This process can be somewhat unconstrained, and might place a large load on working memory. In contrast, such generation is not necessary during classification because a full stimulus is presented on every trial, which might enable subjects to better use the compositional hint to discover the corresponding category rule. In line with this proposal, subjects in Experiment 2 were able to use the compositional hint to learn on the classification task, but not on the inference task. However, in the present study the compositional hint seems to have helped subjects' performance on both classification and inference. Engaging in classification before inference might therefore help subjects to better represent the structure of a relational concept, which can in turn constrain the types of inferences that subjects subsequently make, thus attenuating the difficulty of using compositional-based representations during inference. Nevertheless, further work will be required to more directly test this possibility.

---

[6] These analyses were based on each condition's difference scores between performance on the classification and inference task.

Representational Preference and Individual Differences
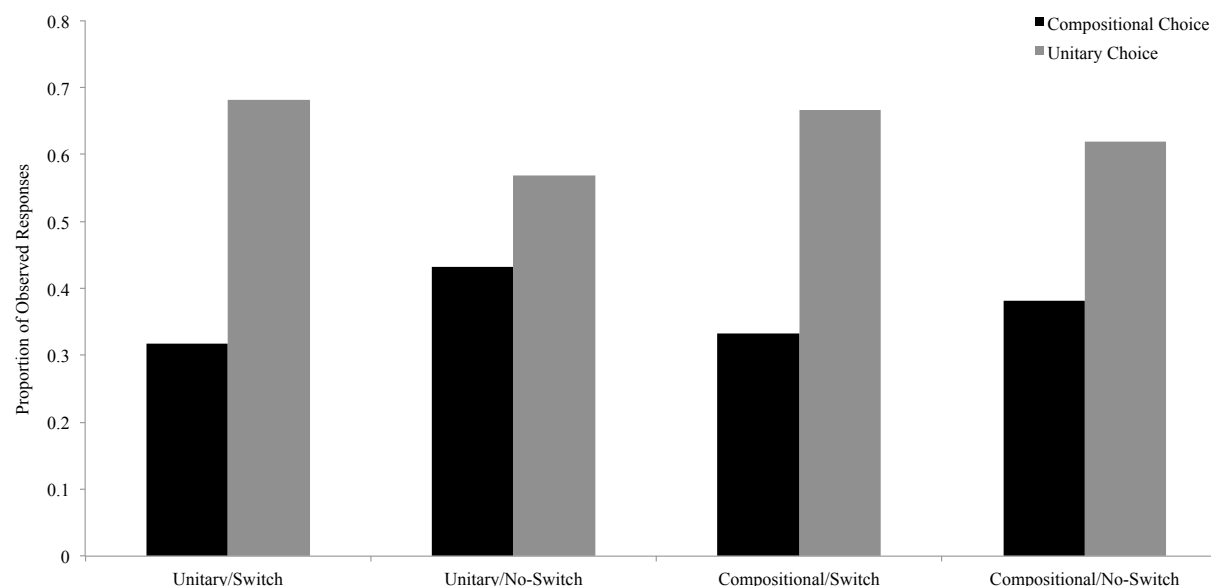
Figure 18. Proportion of observed responses of each category rule in each condition in Experiment 4.

Figure 18 shows the proportion of observed responses in each condition. As in Experiments 2 and 3, a binomial test (collapsing across all conditions) revealed that subjects selected the category rules that were described unitarily at a higher proportion ($M = .62$) than those that were described compositionally ($M = .38$), $p = .001$. However, a subsequent logistic regression showed no differences among the conditions in the type of category rules that subjects selected, all ps $= \geq .27$. Additionally, no relationship was found between the type of category rule that subjects selected and performance on the classification and inference task, both ps $> .50$. Nevertheless, the initial result reported here provides further support for the findings from Experiments 1-3, which suggest that subjects represent relationally structured stimuli unitarily.

An additional analysis was conducted to further explore the idea that there are individual differences in how subjects prefer to represent relational stimuli. Experiment 2 showed that for

subjects who received a unitary hint on the inference task, those who selected the category rules that were described compositionally outperformed subjects who selected the category rules that were described unitarily. This same comparison was conducted in the current study (to replicate the previous finding from Experiment 2) and the results from Experiment 2 were replicated, as subjects in the unitary/no-switch condition who selected the compositional-based category rules ($M = .756$) outperformed (on the inference task) subjects in this same condition (i.e., unitary/no-switch) who selected the unitary-based category rules ($M = .651$), $t(42) = 2.94$, $p = .005$, $SE = .018$, $d = .91$. These findings can be interpreted as providing further support for the idea that subjects who are more inclined or better at representing relational concepts compositionally are better able to use a unitary representation to construct a structured concept than subjects who are more inclined to represent such concepts unitarily. Thus, although subjects might prefer to represent relational stimuli unitarily, there might be a benefit to representing or learning structured concepts compositionally. However, it is important to note that this finding is correlational and this interpretation is speculative. As such, there are other possible interpretations that might account for this result. For instance, it is possible that subjects who have higher working memory capacity also prefer to represent or are better able to learn the categories compositionally, and thus better learn the task than subjects who prefer to represent or learn the categories unitarily. Thus, the interpretation for this finding, although plausible, should be taken with caution.

Summary of Experiments 2-4

Experiments 2-4 examine how unitary and compositional hints, which are meant to encourage corresponding representations in subjects, affect relational learning. First, it is important to note that all three studies provide support for the findings from Experiment 1, which

suggests that people typically represent relational concepts unitarily, but there also appear to be individual differences in how such concepts are represented. The findings from Experiment 2 showed that both unitary and compositional hints can aid learning on a classification task, but only the unitary hint was a useful learning aid on the inference task. These findings provide support for the idea that subjects can indeed use both types of representations to understand and learn relational concepts, but that unitary representations are as or more effective than compositional ones. This latter conclusion challenges the emphasis on compositional representations at the core of most research on analogical reasoning.

Experiment 3 used only a classification task and was able to replicate the findings from the classification condition in Experiment 2, as subjects who received only a unitary hint outperformed subjects who received only a compositional hint. Furthermore, the results from this study showed that subjects who received a unitary hint at any point in the study (with a compositional hint coming before, after, or not at all) outperformed subjects who did not receive a unitary hint at all. No differences in performance were found among subjects in the groups who received a unitary hint. These results lend further support to the dominance of unitary representations, in that subjects will abandon or ignore suggestions for compositional representations if they have discovered a unitary one.

Experiment 4 extended the study design of Experiment 3 and introduced an inference task after classification. Although no differences were found among the groups on the inference task, subjects who started out using a compositional hint and were then presented the unitary hint showed a statistically reliable improvement in performance on the inference task (no other groups showed such improvement) and a greater improvement than subjects who started out the study with a unitary hint. For subjects who begin the study using a compositional hint, one

benefit to engaging in classification before inference is that classification might help subjects form a more mature representation of the corresponding relational concepts, which they can then use to constrain the type of relationships they consider among a stimulus's components during inference. These findings point to the possibility that in order for compositional representations to be useful during inference, subjects must hold the appropriate requisite knowledge about the corresponding concepts. Nevertheless, once such knowledge has been acquired, it seems that both types of representations can be equally effective.

CHAPTER VI

Experiment 5a

This study builds on the findings from Experiments 1-4 and is premised on the assumption that relational concepts can indeed be represented in two fundamentally different ways (i.e., unitarily and compositionally). The present study examines how different types of learning tasks, specifically classification and inference, interact with different types of comparisons (within- vs. between-category), to affect representation and relational learning. These two variables (i.e., type of task and comparison) are included in the present study because both can be used to test predictions that follow from a given set of representational assumptions.

A seemingly straightforward prediction that follows from structure-mapping theory (Gentner, 1983) is that learning should be best when items from the same relational category are compared, as this allows for their common structures to be more easily aligned and abstracted (Gentner, 1983; Lassaline, 1996; Lassaline & Murphy, 1998), that is assuming that relational concepts are represented compositionally. In contrast, if stimuli are represented unitarily and are encoded similarly to features, comparing items from different categories should lead to superior learning, as this type of comparison highlights the discriminative attributes between the two categories (as predicted by theories of attention; Kruschke, 1992; Nosofsky, 1986).

Corral et al. (2017) tested these predictions using a paradigm in which subjects were shown side-by-side co-presented items (i.e., two-item trials) that were either in the same (within-category comparison) or a different category (between-category comparison; this was a between-subjects study); on every 5th trial, subjects were shown a one-item test trial. In contrast to the prediction that follows from structure-mapping theory (Gentner, 1983), Corral et al. found that subjects who compare items from different categories perform better on both feature- and

relation-based categories. Moreover, the contrast advantage for relation-based categories was shown across four experiments using various stimuli, which varied from instantiating simple perceptual relations to more complex, abstract relations. On the surface, these findings appear to support the idea that structured concepts are not represented compositionally, and thus pose a challenge for theories of relational learning that are premised on the assumption that relational concepts are represented compositionally (e.g., structure-mapping theory). However, Corral et al. offer an alternative explanation of their findings that is somewhat compatible with a structure-mapping framework. In cases where the structure for stimuli from different categories can be partially aligned, it is possible that attention is drawn to the differences in the relational properties between the stimuli, similarly to the manner in which people readily notice the featural differences among items that are analogous when their common structures are aligned, as is the case with *alignable differences* (Gentner & Markman, 1994). This latter proposal relies on the assumption that for some yet unknown reason, attending to the distinctive relational properties between two categories leads to superior learning than from mapping and abstracting a complete structure. One possibility is that these distinctive relational properties are treated as features and are used to identify each category, thus placing less strain on working memory than representing each category's relational structure (e.g., representing the Morkel category through the relation of *functions* as opposed to representing its relational structure).

It is important to point out that the idea that people learn through impediments to alignment is premised on the assumption that relational concepts are represented compositionally, as it is the analogous components between two stimuli that are mapped. If it is the case that the contrast advantage reported by Corral et al. (2017) arose from partial alignment, then a contrast advantage might be expected on both a classification and an inference task, as

subjects should be able to partially align co-presented items and recognize their distinctive relations. However, if the contrast advantage was due to subjects representing the stimuli unitarily, an interaction might be expected between type of comparison and task type. To further elaborate on this prediction, if classification judgments can be made based on the presence of a unitary property, without learning the stimuli's relational structure, then (as described above) that property should be highlighted by contrasting stimuli. On the other hand, making an accurate inference about a relational stimulus requires learning the relationship among the stimulus' component parts, and thus the stimulus must be represented compositionally. As such, a match advantage should be expected on an inference task, because subjects should be able to align the elements and abstract the common structure between two analogous items, as predicted by structure-mapping theory (Gentner, 1983), whereas full alignment is not possible in the contrast condition. As a brief aside, the prediction that inference relies on compositional representation might seem antithetical to the findings from Experiment 2, which showed that a unitary hint leads to better inference learning than a compositional hint. However, if the reader will recall, the explanation for this finding was that a unitary hint aids subjects in building a structured representation, which is indeed represented compositionally.

To briefly summarize, the impediments to alignment hypothesis predicts a match advantage on both the inference and classification tasks, whereas an interaction is predicted if subjects' representation of the categories varies by task. Specifically, if subjects represent the categories compositionally during classification a contrast advantage would be expected, whereas a match advantage would be expected on the inference task if the categories were represented unitarily. These predictions are tested in the present study, which applies the paradigm from Corral et al. (2017). Additionally, half of the subjects completed a classification

task and the other half completed an inference task. The stimuli from Experiments 2-4 were used in the present study, but were slightly modified (as described further below). After each response, subjects were given corrective feedback (i.e., shown whether their response was correct and shown the correct response).

Method

Participants

Two hundred ninety-three subjects participated in this study for course credit in an introductory psychology course at the University of Colorado Boulder. Subjects were randomly assigned to four conditions: contrast/classification (N = 73), match/classification (N = 78), contrast/inference (N = 72), and match/inference (N = 69).

Stimuli and Design

This was a 2 (classification vs. inference) × 2 (match vs. contrast) between-subjects experiment with performance on one-item trials as the primary dependent measure. One-item trials were used as the primary dependent measure to control for potential differences that might arise due to the differences in the two types of comparison (e.g., one type of comparison is more difficult to make than the other). The stimuli from Experiments 2-4 were used in this study, but were modified to ensure that none of the components from one machine shared any secondary relations with the components from a separate machine.[7]

---

[7] This modification was made to accommodate an initial plan to randomly sample inference lures for Morkel items from within the Morkel stimuli. This modification would ensure that only one option could be correct on inference trials that consisted of a Morkel stimulus.

Procedure

Subjects completed this study on an LCD 16-inch computer monitor and entered their responses using a computer keyboard. All stimuli were presented on a black background. Subjects were told a cover story that they were going to be shown descriptions of machines made by one of two alien species: Morkels and non-Morkels (this was an A/¬A classification task, as in Experiments 2-4). Subjects were shown a positive example of a Morkel and were instructed to press the spacebar when they were ready to begin the study.

Subjects in all conditions completed 160 trials, which consisted of one- and two-item trials. Two-item trials consisted of two side-by-side items. In the match condition (within-category comparison), both items were from the same category. In the contrast condition (between-category comparison), both items were from different categories. One-item trials were presented every fifth trial and only consisted of a single stimulus, for which subjects were either required to make a classification judgment or an inference (as in Experiments 2 and 4). On one-item classification trials, a stimulus was presented at the center of the screen and subjects were asked to type "A" if the machine was a Morkel or "L" if it was not; subjects in the inference condition were told to type "A" if the top choice was the correct response or "L" if the bottom choice was the correct response.

The classification condition consisted of an A/¬A task. In this condition, the machine that was presented on the left was labeled "Machine A" and the machine that was presented on

---

However, for logistic reasons, these plans were changed and the inference lures for each stimulus (both Morkels and non-Morkels) were instead selected by the researcher, as in Experiments 2-4. The modified stimuli were nevertheless kept in place.

the right was labeled "Machine B"; these labels were presented directly above each machine as shown in Figure 19. On two-item trials, subjects in the match/classification condition were asked to type "2" if both machines were Morkels or "0" if neither machine was a Morkel, whereas subjects in the contrast/classification condition were asked to type "2" if Machine A was a Morkel and Machine B was not or "0" if Machine B was a Morkel and the Machine A was not.
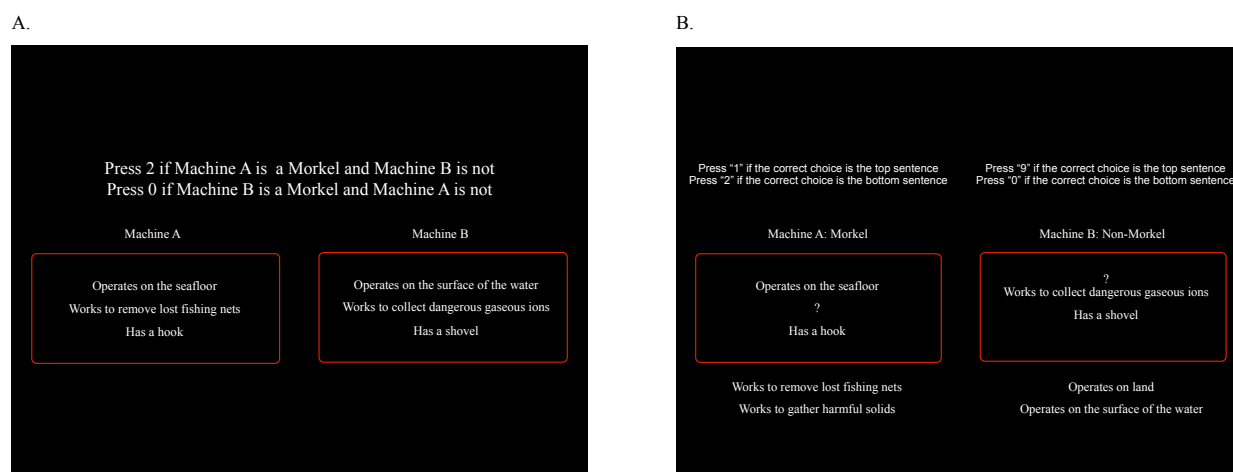
Figure 19. Example trials from the contrast conditions in Experiment 5. A: Example trial from the classification condition. B: Example trial from the inference condition.

In the inference condition, each stimulus was presented with its corresponding category label and two of its three sentences. For each stimulus, subjects were required to infer the missing component (i.e., sentence) from two options, which were presented below each item (as in Experiments 2 and 4). Figure 19 shows example stimuli from a two-item trial for the classification and inference conditions. Response prompts were presented above each machine's label. For Machine A, the prompt instructed subjects to press "1" if the correct choice was the top sentence or "2" if the correct choice was the bottom sentence; for Machine B, the prompt instructed subjects to press "9" if the correct choice was the top sentence or "0" if the correct

choice was the bottom sentence. On two-item trials, each inference option that a subject selected was underlined. Subjects could change their responses by entering the other key press option, which would then show the corresponding inference option underlined. For each stimulus, only the response that was last entered was underlined. On each two-item trial subjects were instructed to finalize their response by pressing the spacebar. Additionally, on each of these trials subjects were required to make an inference for each stimulus in order to move on to the next trial.

As in Experiments 2 and 4, the spatial position of each inference lure was randomized on each trial. The component that subjects were asked to infer was also randomized on each trial, subject to the constraint that subjects were not asked to infer the same type of component for two machines on the same trial. Each inference stimulus had two possible inference lures, one of which was randomly selected for the given trial. For Morkel items, each inference lure shared no secondary relations with any of the machine's components. For non-Morkel items, each inference lure shared one secondary relation with one of the machine's components. For example, as shown in Figure 19B, the correct inference for the non-Morkel stimulus is *operates on the surface of water*, because a shovel (the stimulus's implement) can be used on land (the inference lure) and thus the two share a secondary relation.

The entire stimulus set was presented over the course of 20 trials and the order in which the stimuli were presented on each block of 20 trials was randomized, subject to the constraint that any co-presented items on two-item trials were not members of the same family set. To remind the reader, stimuli were created in sets of families. For instance, one family set of stimuli came from Rehder and Ross (2001), who created three Morkel items and three non-Morkel items by shuffling the features of the Morkel items, in a way that each non-Morkel item took one

feature from each of the three Morkel items; Higgins (2012) and Corral at al. (2017) used this same procedure to each create their own family set of stimuli. This constraint was put in place to ensure that subjects in the contrast/inference condition would not have to make an inference about a component for one machine that was present in the other machine on the same trial.

Subjects were provided corrective feedback on two-item trials, but not on one-item trials; no feedback was provided on one-item trials because these were intended to be test trials. Feedback was presented directly under each stimulus and remained on the screen for 3 seconds. For each stimulus in the inference conditions, both options remained on the screen and the correct option was shown in green and the incorrect option was shown in red. On one-item trials, the stimulus was presented at the center of the screen. After each response (on one-item trials), the screen was cleared and subjects were shown a thank you prompt, which was presented at the center of the screen for 500 ms. The intertrial interval was 400 ms. Subjects were given self-paced rest breaks every 20 trials, in which they were shown their percentage of correct responses over those trials (i.e., one- and two-item trials) and were shown the number of trials they had completed along with the number of trials that remained in the study. Subjects were instructed to press the spacebar when they were ready to continue. The study was set to run for 55 minutes, but subjects who went over the allotted time were allowed to finish the study.

Results and Discussion

A.

B.

Figure 20. Mean learning curves on one- and two-item trials and standard errors of the mean for each condition across blocks of 20 trials in Experiment 5a. A. One-item trials. Each data point represents an average over four one-item trials. B. Two-item trials. Each data point represents an average over 16 two-item trials.

Figure 20 shows learning curves for each condition on one- and two-item trials. One subject was removed from the analysis for scoring over 2 standard deviations below the mean. The results show a marginal interaction on one-item trials between type of comparison and task type, $F(1, 288) = 3.70$, $p = .055$, $MSE = .027$, such that there is a match advantage for the inference task ($M_{contrast} = .64$, $M_{match} = .70$, $t(139) = 2.20$, $p = .029$, $SE = .013$, $d = .373$) and slight (not statistically reliable) contrast advantage for the classification task ($M_{contrast} = .727$, $M_{match} = .711$, $t(149) = .57$, $p = .573$, $SE = .014$). This finding is in line with the idea that subjects represent stimuli unitarily on a classification task, but compositionally on an inference task.

However, the contrast advantage on the classification task found by Corral et al. was not statistically reliable in the present study. It is important to note that although the stimuli used in the present study were taken from Experiment 4 of Corral et al. (2017), these stimuli were modified and a restriction was put in place so that co-presented items were not members of the same family set. Moreover, subjects in the study by Corral et al. completed 300 classification trials, but only 160 in the current study. Nevertheless, a numerical difference was observed in the predicted direction, which qualitavitely replicates the pattern observed by Corral et al. It is also worth noting that the contrast advatange does increase as as the study progresses, as there is a noticable increase in the difference between the contrast and match classification conditions at around the 100th trial, as shown in Figure 20. One possibility is that aforementioned changes weakened the contrast advatnage, but that it is still present and can be observed on later trials.

Experiment 5b

Although Experiment 5a showed the predicted interaction, the contrast advantage was considerably weaker from what was reported by Corral et al. (2017). Experiment 5a was

therefore re-run with subjects in the classification condition completing 300 trials and subjects in the inference condition still completing 160 trials. The two task types consisted of a different number of trials because the study was designed to run for 55 minutes and inference trials take longer to complete than classification trials. Thus, the inference task consisted of fewer trial numbers than the classification task. Two hundred seventy-seven subjects were randomly assigned to four conditions: contrast/classification (N = 70), match/classification (N = 70), contrast/inference (N = 66), and match/inference (N = 71).

Results and Discussion

Because Experiment 5a showed no statistically reliable differences in classification performance between the contrast and match conditions on 160 trials, to reduce the chance of a type two error only classification trials 165-300 (one-item trials) were included for analysis in the present study. First, the results once again revealed a statistically reliable interaction, $F(1, 273) = 5.07$, $p = .025$, $MSE = .032$.[8] Once again, there was a contrast advantage (non-significant trend) for the classification task ($M_{contrast}$ = .85, M_{match} = .803, $t(138) = 1.46$, $p = .146$, $SE = .016$, $d = .25$)[9] and a match advantage (marginal) for the inference task ($M_{contrast}$ = .654, M_{match} = .705,

---

[8] The interaction is marginally significant when all one-item classification trials are included in the analysis, $F(1, 273) = 2.73$, $p < .10$, $MSE = .028$.

[9] This is a marginal effect using a one-tailed test ($p = .073$), which is appropriate to use given that there is an a priori prediction and the classification condition is a replication attempt of two prior studies. It is argued here that for replications, not only is it appropriate to use a one-tailed test, but this practice should be the norm, as honest efforts to replicate a study should take extra precautions to guard against type two errors, particularly in cases where the original result was found using a two-tailed test.

$t(135) = 1.74$, $p = .084$, $SE = .015$, $d = .30$; one-tailed $p = .042$). Although the contrast advantage on the classification task on one-item trials is not statistically reliable, it is when one- and two-item trials are combined ($M_{contrast} = .867$, $M_{match} = .803$, $t(138) = 2.04$, $p = .043$, $SE = .016$, $d = .347$; $p = .02$ with a one-tailed test).[10] Figure 21 shows learning curves for each condition on one- and two-item trials.

The present findings replicate Experiment 5a, however the results could be stronger. One issue to consider is that the present results are likely underpowered. Although making the classification study 300 trials seems to have produced a stronger effect size for the slight contrast advantage than was observed in Experiment 5a, it is still relatively weak. To remedy this issue, follow-up studies should seek to double the trial number on the classification task and increase the sample size; a power analysis (power = .80, $d = .25$) indicates 253 subjects are required per condition to show a statistically reliable effect. However, given that the effect size appears to be stronger on later trials, fewer subjects might be required if the trial number on the classification task is increased.

---

[10] No interaction was found between trial and comparison type ($p > .05$), and thus the two trial types can be treated equivalently and combined, as discussed in Corral et al. (2017).

A.
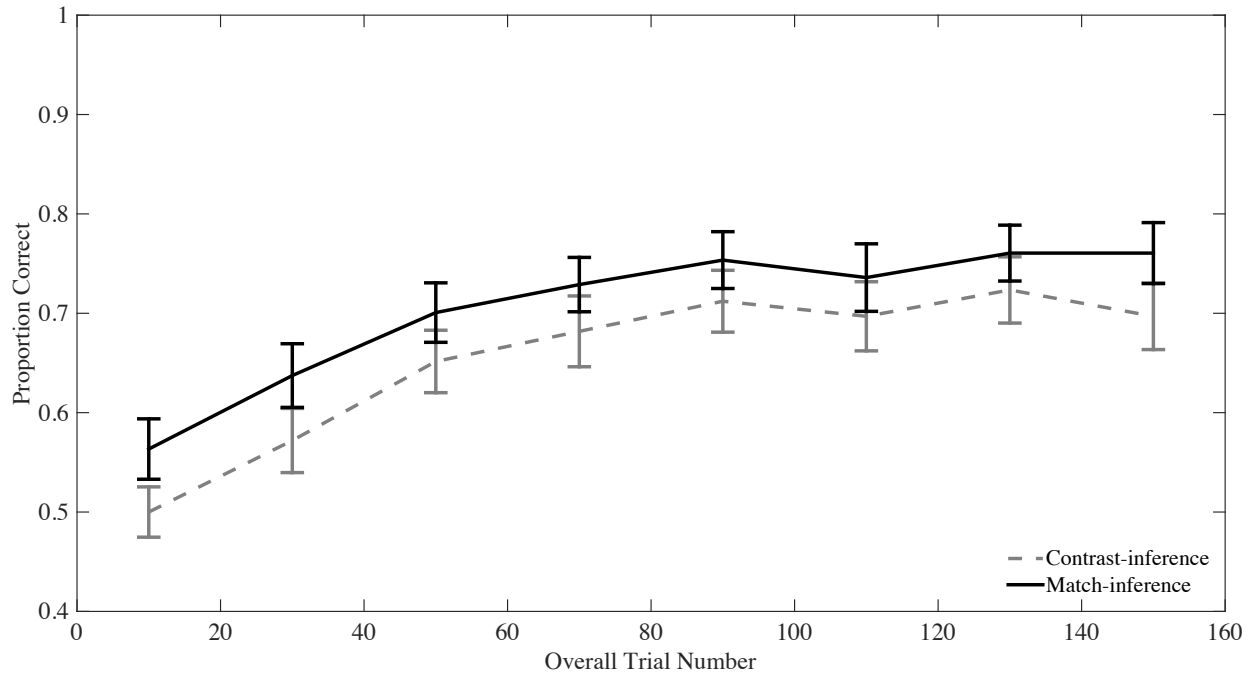
B.
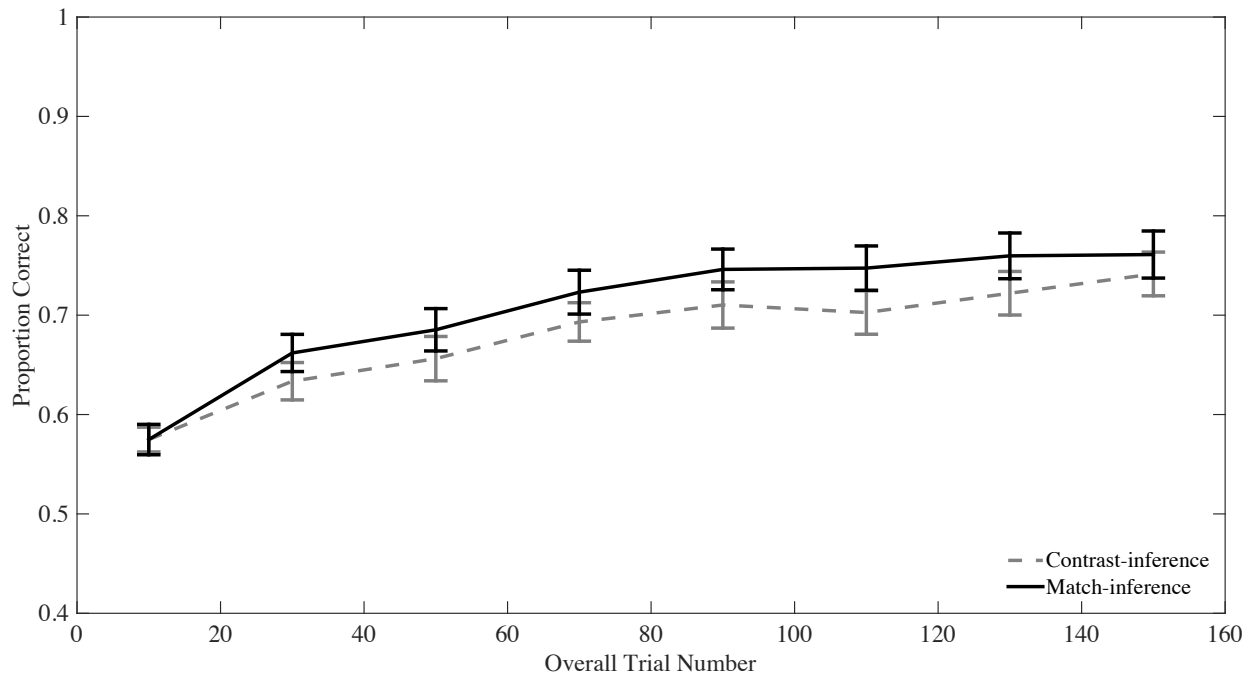
C.
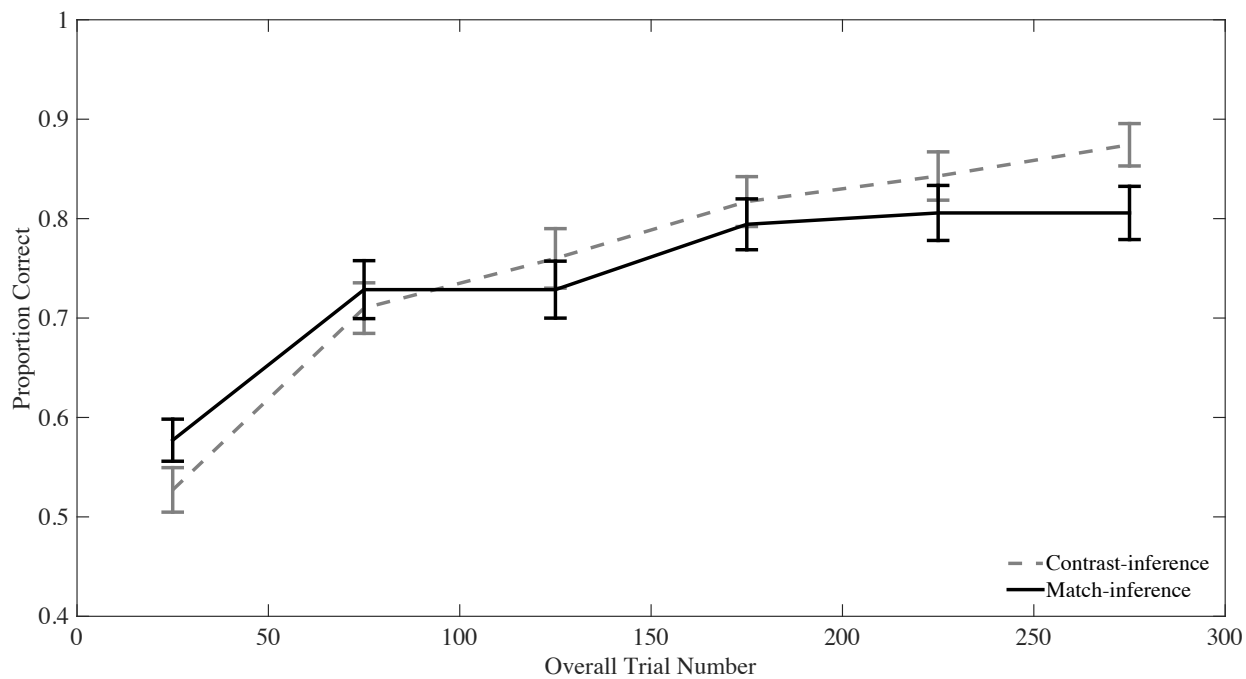
D.

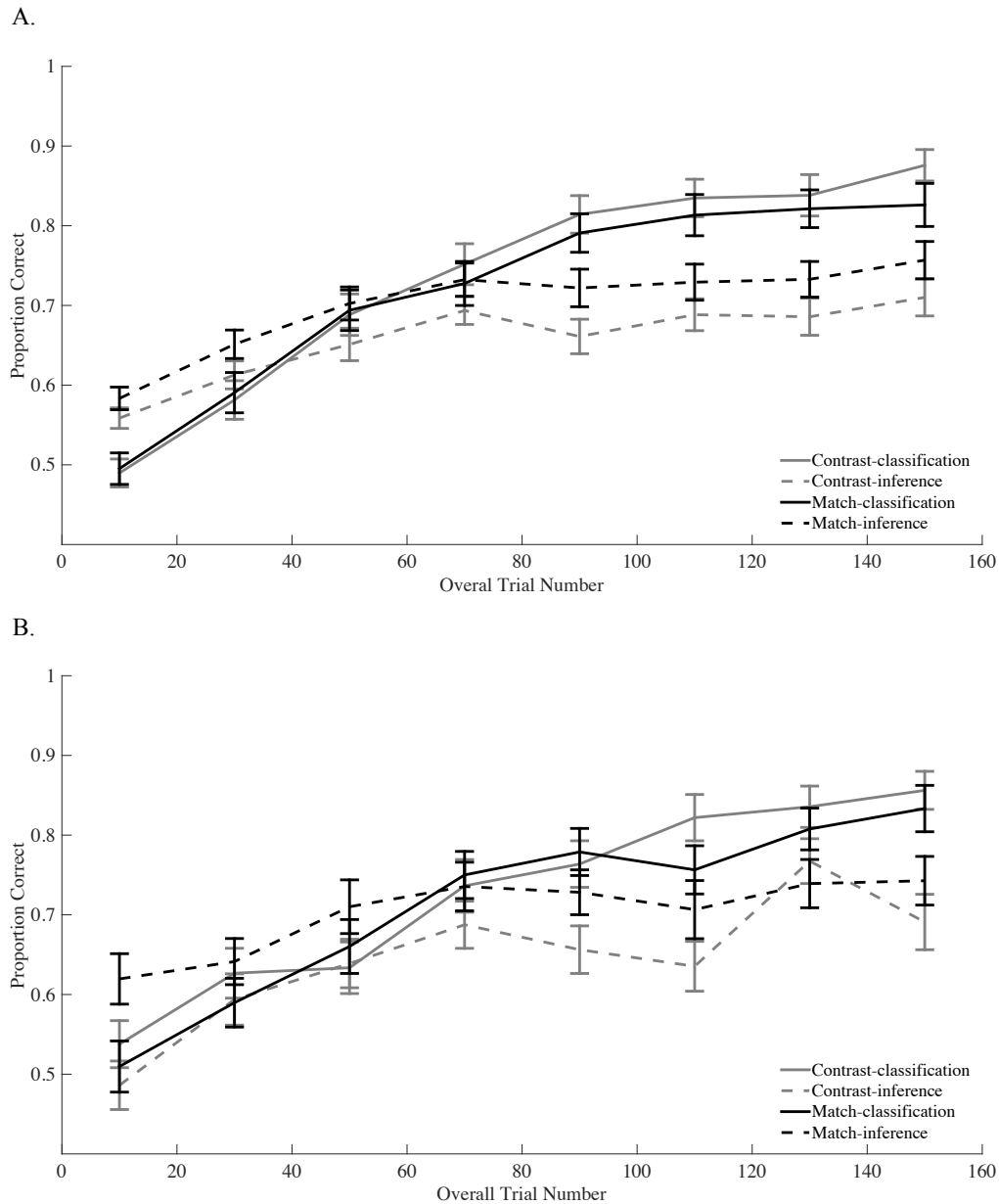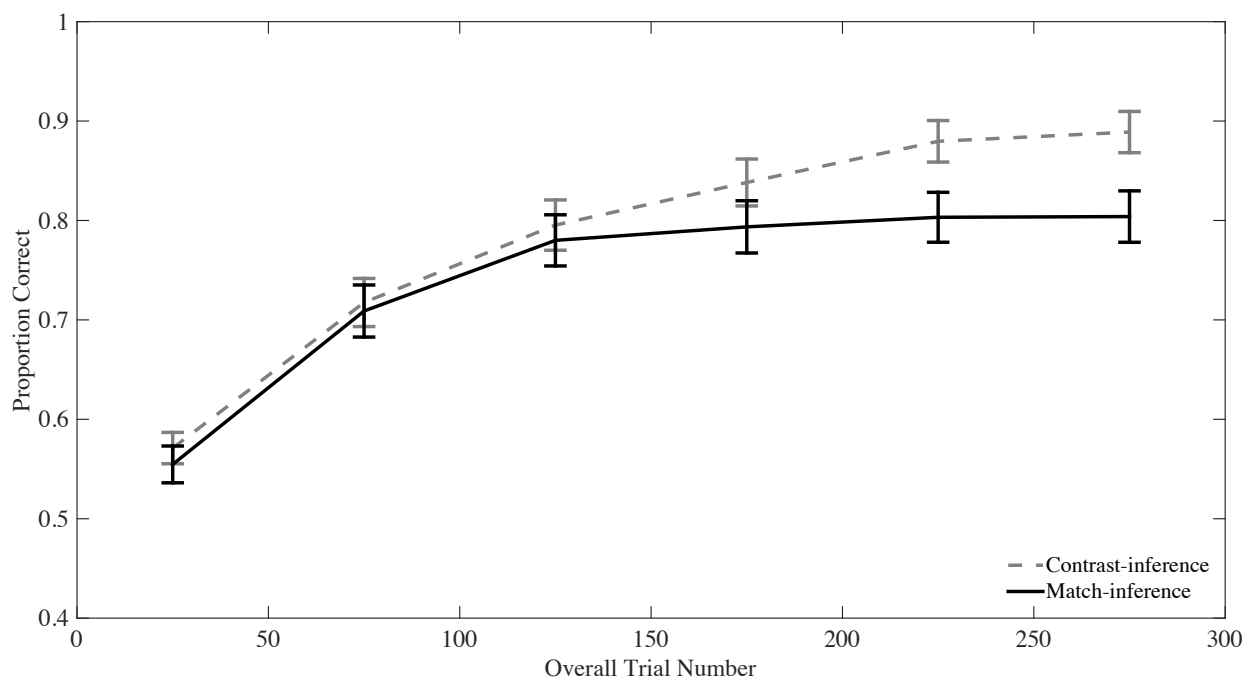*Figure 21*. Mean learning curves on one- and two-item trials and standard errors of the mean for each condition across blocks of 20 (A-B) and 50 (C-D) trials in Experiment 5b. A. One-item

trials for the inference conditions. Each data point represents an average over four one-item trials. B. Two-item trials for the inference conditions. Each data point represents an average over 16 two-item trials. C. One-item trials for the classification conditions. Each data point represents an average over 10 one-item trials. D. Two-item trials for the classification conditions. Each data point represents an average over 40 two-item trials.

Experiments 5a and 5b produced similar results, and showed a slight contrast advantage for the classification task and a match advantage for the inference task. Nevertheless, to ensure that such findings can be trusted, the data for both experiments 5a and 5b were combined to increase statistical power. The results showed a reliable interaction (including all one-item trials from Experiment 5b), $F(1, 565) = 6.53$, $p = .011$, $MSE = .027$, as there was a non-significant contrast advantage in the classification condition ($M_{contrast} = .743$, $M_{match} = .725$, $t(289) = .82$, $p = .42$, $SE = .01$)[11] and a reliable match advantage in the inference condition ($M_{contrast} = .649$, $M_{match} = .704$, $t(276) = 2.80$, $p = .006$, $SE = .01$, $d = .34$).

In sum, although the findings from Experiments 5a and 5b are not definitive and the effect sizes are not particularly large, the findings are nevertheless reliable. The interaction provides support for the idea that relational classification and inference tasks lend themselves to different types of representations. Specifically, the match advantage on the inference task suggests that an inference task requires subjects to represent relational categories compositionally, as a unitary representation might not be readily accessible given that the

---

[11] The contrast advantage is statistically reliable (including all one-item trials; $M_{match} = .75$, $M_{contrast} = .79$) when the data from Experiment 5b are combined with the data from Experiment 4 in Corral et al. (2017), $t(265) = 2.07$, $p = .039$, $SE = .01$, $d = .254$.

stimulus is incomplete. Moreover, the comparison task specifically required subjects to compare the component parts between two machines and discover how two corresponding components between two machines were analogous. This type of process seems to require representing the categories compositionally, and indeed, the match advantage is in accord with the prediction that follows from structure-mapping theory (Gentner, 1983), as comparing analogous items should facilitate aligning their corresponding elements and abstracting their shared structure. On the other hand, the contrast advantage in the classification task suggests that when possible, subjects will represent relational concepts unitarily, as stimuli, particularly those used in the present study, can be classified without the need to represent their structure (as demonstrated in Experiments 2-4). As discussed above, in cases where a concept can be represented unitarily, comparing items from different categories should highlight their distinctive properties (because a unitary concept can treated similarly to a feature) and lead to better learning than comparing items from the same category.

CHAPTER VII

Experiment 6

One possibility is that the interaction found between task and comparison type in Experiments 5a and 5b itself depends on the type of relational stimuli that subjects are presented. More specifically, the stimuli used in Experiments 2-5 might have been well-suited to be represented unitarily, due to the secondary relations among each machine's component parts, which might have activated subjects' pre-existing general knowledge structures about machines, which could in turn be used to find a unitized property within each stimulus (e.g., the description of the machine *makes sense*). One possibility is that in cases when such representations are not readily available, subjects are forced to represent relational concepts compositionally. In such cases, a main effect of comparison might be found, such that there is a match advantage on both inference and classification, as the only way to learn the categories might be to align the component parts between co-presented items. This study tests this prediction and uses the same paradigm as in Experiment 5 (i.e., 2 (match vs. contrast) × 2 (classification vs. inference) between-subjects study with one-and two item trials), along with stimuli that were designed with the specific purpose of inhibiting a unitary-based representation from readily emerging.

Specifically, a stimulus consisted of three terms that pertained to perceptible objects (e.g., *train*, *hamster*, *chocolate bar*), which were arranged vertically and were bounded by a red border (as shown in Figure 22). The stimuli were divided into two categories, defined by the specific manner in which three objects were related to one another based on size. For one category (Zorpes), the third object was *bigger than* the second and the second object was *bigger than* the first object; for the other category (Olatin), the first object was *bigger than* the second and the second object was *bigger than* the third object. Figure 22 shows example stimuli from the

contrast/classification condition in Experiment 6. Appendix C contains the stimuli that were used in this study.
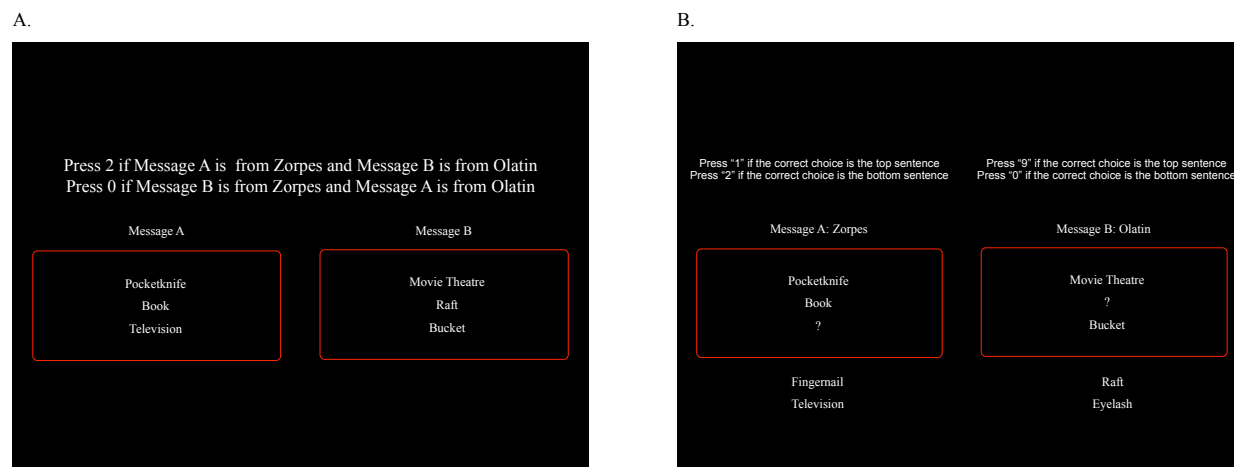
Figure 22. Example trials from the contrast conditions in Experiment 6. A. Classification condition. B. Inference condition.

The stimuli in Experiments 2-4 leveraged subjects' prior knowledge about the secondary relationships among a stimulus' component parts. Corral et al. (2017) posit that subjects' knowledge about these secondary relations might have given rise to a unitized representation. However, the stimuli in the present study lacked this property in that the objects within any given stimulus are not typically associated with one another and do not typically share a common category (at least not on the relation that defined the categories), which might therefore inhibit the emergence of a unitized representation. For instance, consider the following example stimulus: *great white shark*, *sunglasses*, *light bulb*. Although there are certainly many relationships among these objects, such relationships are somewhat obscure and might not be readily salient to subjects. As a result, in order to learn the category rule for these stimuli it might be necessary for subjects to represent the component parts of each stimulus (i.e., each object) and

explicitly consider how they are related to one another, a process that would likely rely on representing the categories compositionally. Due to subjects shifting their attention to the stimuli's component parts, the emergence of a unitary concept might be disrupted. Nevertheless, it is important to note that each category rule could be represented unitarily. For Zorpes, the category rule could be thought of as an instantiation of the concept of *grow*, in that the objects in a stimulus grow or increase in size from top to bottom. For Olatin, the category rule could be thought of as an instantiation of the concept of *shrink*, in that the objects in a stimulus shrink or decrease in size from top to bottom. However, to discover such representations subjects would still likely need to explicitly represent the relationships among the component parts within a stimulus (as just explained), as such a representation might only emerge after a subject has discovered how the objects are related, and thus a match advantage should still be expected.

Method

Participants

One hundred ninety-six subjects participated in this study for course credit in an introductory psychology course at the University of Colorado Boulder. Subjects were randomly assigned to four conditions: contrast/classification ($N = 47$), match/classification ($N = 48$), contrast/inference ($N = 50$), and match/inference ($N = 51$).

Stimuli, Design, and Procedure

Thirty-six stimuli were created, 18 for each category. Subjects were provided the cover story that aliens from two different planets were communicating with Earth by sending coded messages. Subjects were told that the two types of messages differed from one another. One type of message was being sent by aliens from the planet Zorpes and the other was being sent by

aliens from the planet Olatin. Subjects were told that it was their job to figure out which type of message was sent by each planet.

Directly above each stimulus on all (classification and inference) two-item trials, the stimulus on the left was labeled "Message A" and the stimulus on the right was labeled "Message B". In the classification condition, subjects completed a two-category classification task. On two-item trials, subjects in the match/classification condition were told to press "2" if both messages were from Zorpes or "9" if both messages were from Olatin, whereas subjects in the contrast/classification condition were told to press "2" if the message on the left was from Zorpes and the message on the right was from Olatin or "9" if the message on the left was from Olatin and the message on the right was from Zorpes. On one-item trials, subjects in the classification condition were instructed to type "A" if the message was from Zorpes or "L" it was from Olatin; subjects in the inference condition were told to type "A" if the top choice was the correct response or "L" if the bottom choice was the correct response. All response prompts were presented above the stimuli at the center of the screen.

For each two-item trial in the inference condition, each category label was presented directly above the stimulus. Two inference lures were created for each object within a given stimulus. On each inference trial, the inference lure was randomly selected for each stimulus. Each inference lure made the stimulus violate its category structure. For instance, consider the Zorpes stimulus in Figure 22B. The lure for this stimulus is *fingernail*, because Zorpes are defined by the 3rd object being larger than the 1st and 2nd objects and the 2nd object being larger than the 1st object, but a fingernail is smaller than both a *book* and a *pocketknife*. Appendix C shows the inference lures that were created for each object within a given stimulus. For two-item inference trials, subjects were presented a prompt directly above each message. For Message A,

subjects were instructed to press "1" if the top choice was correct or "2" if the bottom choice was correct; for Message B, subjects were instructed to press "9" if the top choice was correct or "0" if the bottom choice was correct.

Subjects in the classification condition completed 200 trials and subjects in the inference condition completed 80 trials; the difference in trials between the two conditions is due to time constraints, as the study was designed to be completed within 30 minutes and the inference task takes longer for subjects to complete than the classification task. At the end of the study, subjects were shown two descriptions of each category rule, one of which was described unitarily and the other was described compositionally. Subjects were asked to select which option best represented how they were thinking of the two category rules by typing "A" or "B"; these two options corresponded to the two types of descriptions that subjects were shown. The unitary option read as follows: "For one type of message, the objects were organized in ascending order according to size, whereas for the other message the objects were organized in descending order according to size". The compositional option read as follows: "For one type of message, the 3rd object was bigger than the 2nd object and the 2nd object was bigger than the 1st object, whereas for the other type of message the 1st object was bigger than the 2nd object and the 2nd object was bigger than the 1st object". The option that corresponded to the two types of descriptions was randomized for each subject (as in Experiments 2-4). The average length of this study was approximately 22 minutes. The rest of the design and procedure were identical to that of Experiment 5.

Results and Discussion

A.

B.
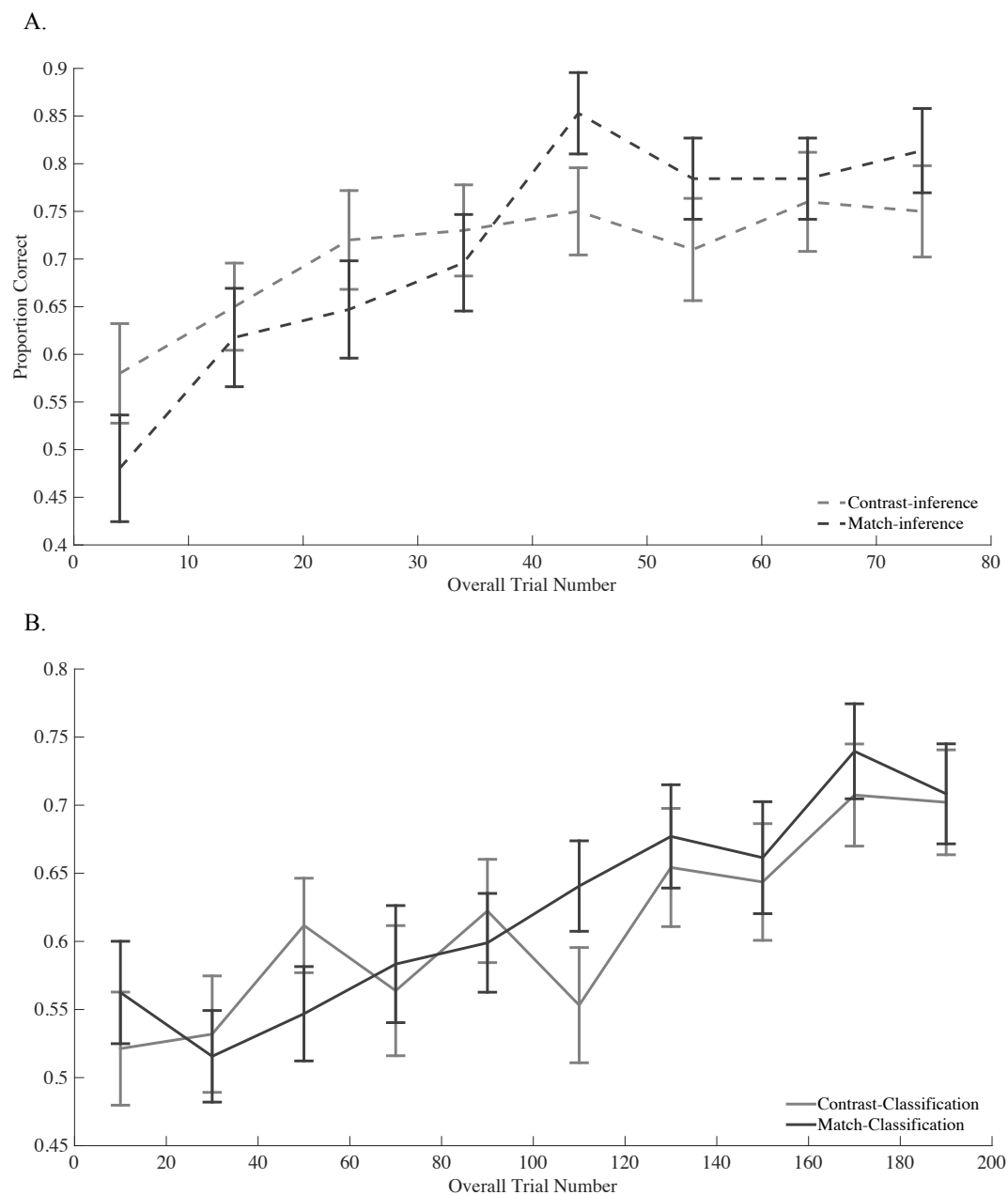
Figure 23. Average learning curves on one-item trials and standard errors of the mean across blocks of 10 (A) and 20 (B) trials for the inference and classification conditions in Experiment 6. A. Inference condition. Each data point represents an average over two one-item trials. B. Classification condition. Each data point represents an average over four one-item trials.

Figure 23 shows average learning curves on one-items trials for subjects in each condition in Experiment 6. A two-way ANOVA revealed no main-effect of comparison, $F(1, 192) = .1$, $p = .755$, $MSE = .03$, and no interaction between comparison and task type, $F(1, 192) = .03$, $p = .858$, $MSE = .03$. Similarly, no differences in performance were found between the match ($M_{classification} = .623$; $M_{inference} = .611$) and contrast ($M_{classification} = .706$; $M_{inference} = .710$) conditions on the classification, $t(93) = .420$, $p = .678$, or inference tasks, $t(99) = .42$, $p = .933$. However, additional analyses do show slightly stronger, albeit non-significant, results. Specifically, when only the second half of the one-item trials were analyzed, there was a non-significant trend for the main effect of comparison, $F(1, 192) = 1.72$, $p = .191$, $MSE = .054$, such that subjects in the match condition ($M_{classification} = .685$; $M_{inference} = .810$) outperformed subjects in the contrast condition ($M_{classification} = .652$; $M_{inference} = .743$) on both the classification, $t(93) = .865$, $p = .390$, and inference task, $t(99) = 1.38$, $p = .171$. These patterns of results are in line with the predicted outcome of a main effect of comparison, such that subjects in the match condition should outperform subjects in the contrast condition on both tasks (because the stimuli were designed to inhibit a unitary representation from emerging). It is also important to note that the match advantage on the inference task (only on the 2nd half of the trials) does replicate the findings from Experiments 5a and 5b. Nevertheless, these findings are not statistically reliable and should thus be interpreted with caution.

One possibility is that the present experiment is underpowered and thus a larger sample size is required to detect statistically reliable effects. Furthermore, the present study was designed to run for 30 minutes (although subjects were allowed more time to finish in cases where it was necessary) and the trials for each task were set based on this time frame. On each task, the second half of trials do show promising results and it is possible that these qualitative

differences between the match and contrast conditions might prove to be statistically reliable if more trials were added to the study (e.g., doubling the trial number on each task). Furthermore, although the stimuli were meant to inhibit unitary representations in subjects, they could be further altered to better accomplish this goal. As mentioned above, it was possible to represent the categories based on whether the sizes of the objects were *ascending* (for the Zorpes category) or *descending* (for the Olatin category). To further inhibit such a representation from emerging, the structure of the Zorpes category could be changed (e.g., the 2nd object is *bigger* than the 1st and the 1st object is *bigger* than the 3rd). The structure of the Olatin category could be altered in a similar manner (e.g., the 3rd object is *bigger* than the 1st and the 1st object is *bigger* than the 2nd object). Category structures such as these might be more challenging to represent unitarily than those used in the present study, as these structures cannot be as easily classified by a readily accessible concept, such *ascension*. Such changes might further encourage subjects to represent the stimuli compositionally (as no other representation might be available). As a result, a greater performance difference between the match and contrast condition might be observed, as within-category comparison is posited to lead to better relational learning than between-category comparison (if stimuli are represented compositionally), because subjects can better align the corresponding elements between two items that share a common structure than when they do not. Nevertheless, further work will be required to better answer these questions.

Representational Preference and Individual Differences

Figure 24 shows the proportion of subjects who selected each of the category rules in each condition. As in Experiments 2-4, a binomial test showed that a greater proportion of subjects selected the category rules that were described unitarily ($M = .783$) than those that were described compositionally ($M = .217$), $p < .0001$. A logistic regression showed no main effects of

comparison or task type and no interaction between these two variables, all ps $\geq$ .19. Furthermore, the type of category rules that subjects selected (collapsing across comparison and task type) did not predict classification performance, $F(1, 194) = 1.39$, $p = .24$, $MSE = .027$.

Follow up exploratory analyses were also conducted to examine whether there were differences in performance within conditions based on the type of category rule that subjects selected. A marginal difference in performance was found between subjects in the match/inference condition, $t(49) = 1.87$, $p = .068$, $SE = .028$, $d = .534$, such that subjects who selected the category rules that were described unitarily (.746) outperformed subjects who selected the category rules that were described compositionally (.636). One possible explanation for this finding is that subjects who were representing the stimuli unitarily were better able to use that representation to structure or constrain the type of inference they made in the match condition. Another possibility is that all subjects in the match/inference condition started off with a compositional-based representation, but that representation evolved into one that was unitary for those who better learned the category structures (i.e., those with higher performance on the inference task). However, these explanations are merely speculative and require further follow-up work.
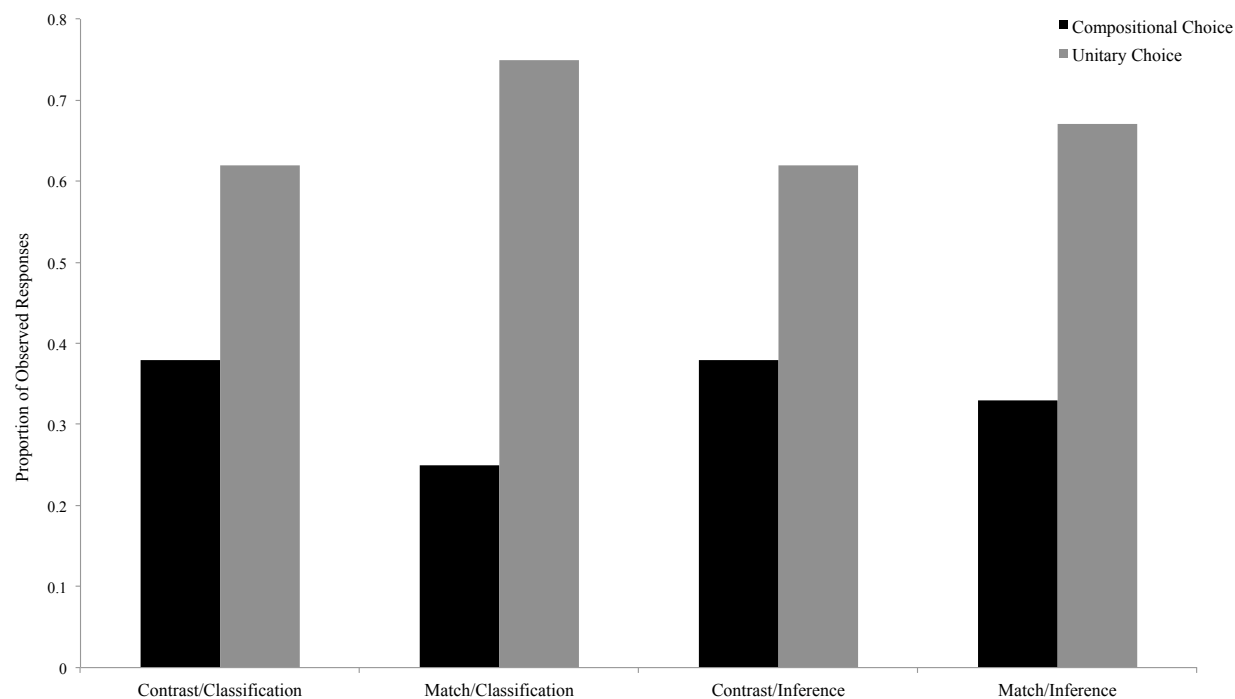
Figure 24. Shows the proportion of subjects who selected each the of category rules in each condition in Experiment 6.

CHAPTER VIII

Experiment 7

Experiment 2 examined the congruence between type of hint and task. Similarly, Experiment 5 examined the congruence between type of comparison and type of task. The present study examines the congruence between type of hint and comparison. More specifically, this study builds on the findings from Experiment 2-6 and examines how using compositional-versus unitary-based hints on a classification task affects learning when subjects compare items from the same versus different categories. To elaborate further, Experiment 2 showed that both compositional- and unitary-based hints improve learning on a relational classification task. Similarly, extensive work has shown that comparison can aid concept acquisition (Alfieri, Nokes-Malach, & Schunn, 2013; Bransford & Schwartz, 1999; Gick & Holyoak, 1983; Schwartz & Bransford, 1998; Ward & Sweller, 1990). For these reasons, the paradigms from Experiments 2 and 6 were partially combined. As in Experiment 6, subjects completed a two-category classification task, wherein subjects completed one- and two-item trials. Additionally, as in Experiments 2-4, subjects were shown a given hint (depending on their condition) every 3rd error. This study was therefore a 2 (match vs. contrast) × 2 (compositional hint vs. unitary hint) between-subjects experiment with performance on one-item trials as the primary dependent measure.

The stimuli in this study were modeled after those used by Foster, Cañas, and Jones (2012), but were modified to fit within the present paradigm. A stimulus consisted of three photo finishes between two spaceships that were racing. There were three spaceships per stimulus and each raced one another once. Thus, Spaceship A raced Spaceship B and C and Spaceship B raced Spaceship C. Figure 25A shows an example of a trial from the contrast condition. The stimuli

were divided into two categories, distinguished by different relational structures. One category was defined by one spaceship beating the other two and thus winning each of its races; the other category was defined by each spaceship winning and losing one of its races. Figure 25B shows the different abstract instantiations of each category's relational structure.

A.
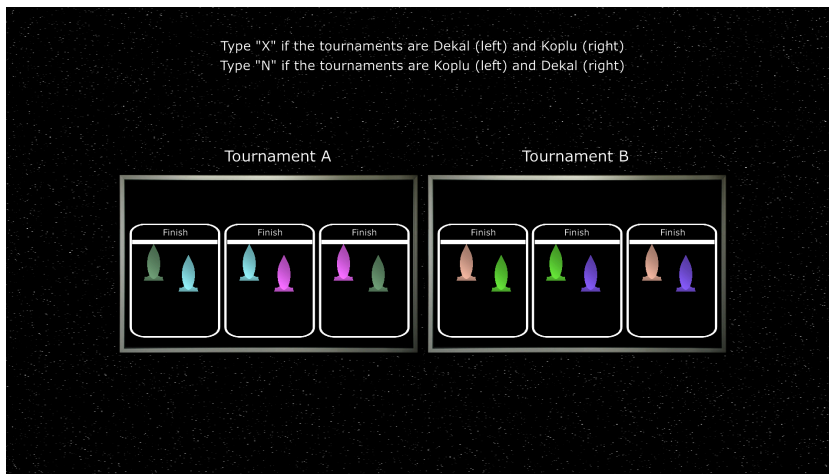
B.

Category A: Cycle   Category B: Hierarchy
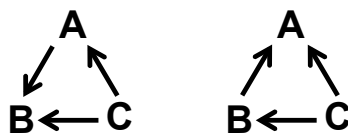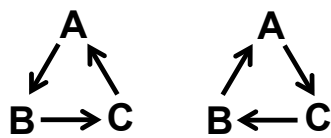
Figure 25. Shows an example trial from the contrast condition in Experiment 7 and the different abstract instantiations of each category structure. A. Example trial from the contrast condition in Experiment 7. The stimulus on the left shows a cycle category, wherein each spaceship wins and loses a race; the stimulus on the right denotes a hierarchy category, such that one spaceship wins all of its races. B. The different abstract instantiations of each category structure. Each letter represents a spaceship and the arrows between two spaceships indicate the outcome of the race between the two, such that arrows pointing away from a spaceship denote the winner of the race and arrows pointing to a spaceship denote the loser (e.g., A ➜ B indicates that Spaceship A beat Spaceship B in a race).

It is important to note that the paradigm used in this study allows for key theoretical predictions that follow from using unitary- versus compositional-based hints to be tested. Theories of attention (Kruschke, 1992; Nosofsky, 1986) predict that comparing feature-based stimuli from different categories should highlight their diagnostic properties. Thus, because unitary-based concepts are posited to be represented similarly to features (Corral et al., 2017), subjects who are shown a unitary hint should perform better in the contrast condition than in the match condition, as the contrasting pairs should highlight the unitary concept that defines each category. A contrasting prediction that follows directly from the findings from Experiment 2 is that subjects will use a unitary hint to construct a compositional-based representation, and thus these subjects will perform better in the match condition than in the contrast condition. A similar set of contrasting predictions can be made for the compositional hint. One straightforward prediction is that a compositional hint will encourage subjects to develop a corresponding representation, which should lead to better performance in the match condition than in the contrast condition. As described above, subjects should be able to align the corresponding

elements and abstract the common structure between items from the same scenario more easily in the match condition than in the contrast condition. Alternatively, as pointed out by Corral et al. (2017), representing the stimuli compositionally might produce a contrast advantage, as the stimuli might be partially aligned, highlighting the contrasting relations between the two categories. Thus, a contrast advantage might be expected for subjects who receive a compositional hint.

Moving on to a different prediction, as discussed above, the compositional hint encourages subjects to look for the interconnections among a stimulus's component parts. This process seemingly relies on bottom-up learning, which can be somewhat unconstrained, as there can be many potential relationships among a stimulus' components. In contrast, a unitary hint can provide top-down structure to a representation, in that the concept itself can constrain the types of relationships that subjects consider among a stimulus' components. This account might partially explain why the unitary hint leads to better learning than the compositional hint. However, in cases where the compositional hint is used within a relatively constrained context, it is possible that both types of hints can be equally effective.

One idea that followed from Experiment 4 was that the advantage of the unitary hint over the compositional hint depends on whether subjects have an additional learning aid that can be used to scaffold learning. It was proposed that the classification condition provided such a constraint, which allowed subjects who received a compositional hint to build up an adequate representation of the category structure, which subjects were then able to transfer and apply on the inference task. In this case, the classification task served as a type of scaffold for relational inference. Experiment 4 thus showed that when such a learning aid is available, a compositional hint can produce similar performance to that of a unitary hint. Although this idea was not directly

tested in Experiment 4, it can be addressed using the present paradigm, as comparison might serve as a similar type of learning aid as the classification task did in Experiment 4, which might thus offset the benefit of the unitary hint over the compositional hint. To elaborate, comparison can draw subjects' attention to the similarities and differences between each stimulus' component parts, which can better highlight (than a single stimulus) their relational structures. Comparison can thus provide additional conceptual constraints to subjects who are using a compositional hint during classification. Thus, an interaction between type of hint and trial type might be expected, such that a unitary hint might be more beneficial to learning on one-item trials, which do not allow for direct comparison between two items, but not on two-item trials, in which subjects can engage in such comparison.

Method

Participants

Two hundred thirty-six subjects participated in this study for course credit in an introductory psychology course at the University of Colorado Boulder. Subjects were randomly assigned to four conditions: unitary/contrast ($N = 59$), unitary/match ($N = 59$), compositional/contrast ($N = 59$), and compositional/match ($N = 59$).

Stimuli and Design

A single stimulus consisted of three side-by-side photo finishes of races, each between two spaceships. Each race was bounded by a white border and had a white finish line. The three races were presented inside of a rectangular box with a dark grey border, as shown in Figure 25A. Each tournament consisted of three spaceships, each with a unique color (e.g., red, blue,

and green). The two spaceships that raced in the first race were randomly selected, as were the spaceships that were presented on the left (Spaceship A) and right side (Spaceship B) of each photo finish. The third spaceship (Spaceship C) raced in the second race against either Spaceship A or B, depending on the instantiation of the corresponding tournament category. There were two possible instantiations of the cycle category and six possible instantiations of the hierarchy category, as shown in Figure 25B. All losing spaceships were presented at the same height in the photo finishes. All winning spaceships were presented at the same height in the photo finishes.

Procedure

All stimuli were presented on an outer space-like background (as shown in Figure 25A) on a 16-inch LCD monitor and all responses were entered using a computer keyboard. As a cover story, subjects were told that they were a sports writer for the Galactic Times and were attending the 875th annual Space Dash. Subjects were also told that there were two types of tournaments, Dekal and Koplu, and it was their job to figure out the difference. Lastly, subjects were notified that they would be shown photo finishes of each type of tournament on two jumbotrons and they would need to be able to distinguish the two.

Before the first trial, subjects were given a hint about the two tournaments, which varied based on the subject's condition. All subjects were shown the hint that corresponded to their condition once more after the 1st trial, on every third error, and on rest breaks (as in Experiments 2-4). Subjects who received a unitary hint were shown the following: "Remember, the two types of tournaments differ by whether their overall outcomes are expected or not. For each tournament, think about whether the outcome of its races is expected or unexpected." Subjects who received the compositional hint were shown the following: "Remember, the two types of tournaments differ by how the ships in each tournament are related. For each tournament, think

about how the 1st spaceship relates to the 2nd and 3rd spaceships, as well as how the 2nd spaceship relates to the 3rd spaceship." On every third error, the screen was cleared and subjects were presented the hint from their condition at the center of the screen and subjects were instructed to stop and read the hint carefully; subjects were instructed to press the spacebar when they were ready to continue.

On each trial, the stimulus category was randomly selected, as was its type of instantiation. Additionally, on each trial, the color of each spaceship was randomly selected from the RGB color spectrum, subject to the constraints that (1) the stimulus was visible on the screen, (2) that no spaceship within a given stimulus was the same color, and (3) that the color of all spaceships within a stimulus were discriminable from one another, such that the difference between each spaceships' summed squared RGB value was greater than 35%.

On two-item trials, subjects in the match condition were told to press "X" if both tournaments were Dekal or "N" if both tournaments were Koplu; subjects in the contrast condition were told to press "X" if the tournament on the left was Dekal and the tournament on the right was Koplu or "N" if the tournament on the left was Koplu and the tournament on the right was Dekal. The labels "Tournament A" and "Tournament B" were displayed directly above the left and right stimulus, respectively. After each response on a two-item trial, subjects were shown whether they were correct along with the correct answer, which was displayed in large white letters inside of each jumbotron, directly above the photo finishes. Feedback remained on the screen for 2 s. On one-item trials, a single stimulus was presented at the center of the screen and subjects were instructed to press "D" if the tournament was Dekal or "K" if it was Koplu. The name of the tournament that corresponded to each category was counterbalanced across subjects within each condition. Subjects were not shown corrective feedback on one-item trials

(because these were test trials). After each response on these trials, the screen was cleared and subjects were shown a thank you prompt for 400 ms. Subjects completed 100 trials.

The intertrial interval was 800 ms. Every 20 trials, subjects were given a self-paced rest break and were shown the proportion of correct responses they answered correctly over those trials, along with the number of trials they completed and the number that remained. Subjects were also shown and reminded to use the given hint they had been using. At the end of the study, subjects were shown two descriptions of the category rules (unitary and compositional, as in Experiments 2-4 and 6) and were asked to select the option that best corresponded to how they were representing the two types of tournaments. The unitary description of the tournaments read as follows: "One tournament was defined by each spaceship winning and losing a race, whereas the other tournament was defined by one spaceship never losing". The compositional description read as follows: "One tournament was defined by Spaceship A beating Spaceship B but losing to Spaceship C, and Spaceship B beating Spaceship C, whereas the other tournament was defined by Spaceship A beating Spaceships B and C". Subjects were asked to press "A" for one option or "B" for the other option. The option that corresponded to the two types of descriptions was randomized for each subject.

Results and Discussion

Seventeen subjects were excluded from the final analysis (unitary/match = 4, unitary/contrast = 2, compositional/match = 2, and compositional/contrast = 6) for having an average response time of less than 1.5 seconds on over 80% of their trials, suggesting that these

subjects did not put forth an honest effort in this study.[12] Table 3 reports the analyses with all subjects included.

Figure 26 shows average learning curves on one- and two-items trials for subjects in each condition in Experiment 7. A two-way ANOVA (restricting to one-item trials, as this was the primary dependent measure) showed a main effect of hint, $F(1, 215) = 4.0$, $p = .047$, $MSE = .036$, such that subjects who received a unitary hint ($M = .82$) outperformed subjects who received a compositional hint ($M = .766$). No effect of comparison was found, $F(1, 215) = 1.32$, $p = .252$, $MSE = .036$, but there was an interaction between type of hint and type of comparison, $F(1, 215) = 6.31$, $p = .013$, $MSE = .036$, indicating that the main effect of hint depends on the type of comparison that subjects engage in. A series of planned comparisons revealed that subjects in the match/compositional condition performed worse than subjects in the other three conditions. Specifically, subjects in the match condition performed better using a unitary hint ($M = .834$) than subjects who used a compositional hint ($M = .721$), $t(108) = 3.11$, $p = .002$, $SE = .018$, $d = .60$. Additionally, subjects in the contrast/compositional condition ($M = .813$) outperformed subjects in the match/compositional condition, $t(108) = 2.46$, $p = .015$, $SE = .018$,

---

[12] The decision to remove these subjects was based on the experimenter's judgment that these reaction times were not sufficient for subjects to adequately process the outcome of two tournaments, which comprise six total races. This screening process was used to remove subjects from the dataset who either did not put forth an honest effort and/or who gave up on the task too early into the study to provide usable data. Nevertheless, it is important to note that the qualitative pattern of the results is unchanged when these subjects are not excluded and the results reported in this section and their interpretations remain the same, as shown in Table 3.

d = .473; subjects in the contrast/unitary condition also outperformed subjects in the match/compositional condition, $t(111) = 2.31$, p = .028, SE = .018, d = .439. No differences in performance were found between subjects in the contrast conditions who used the unitary (M = .804) and compositional hints (M = .813), $t(107)$ = .262, p = .794, SE = .018. Similarly, no differences in performance were found for subjects who received a unitary hint between the match (M = .834) and contrast conditions (M = .804), $t(107)$ = .855, p = .395, SE = .018.

Table 3. Experiment 7 results with all subjects included in the dataset.

	df	t-statistic	p	d	C/U (M)	C/C (M)	M/U (M)	M/C (M)
C/U vs. C/C	116	.36	.72	.07	.792	.778	.80	.714
C/U vs. M/U	116	.16	.873	.03				
C/U vs. M/C	116	2.12	.036	.394				
C/C vs. M/U	116	.51	.611	.095				
C/C vs. M/C	116	1.70	.094	.316				
M/U vs. M/C	116	2.25	.027	.42				

ANOVA (One-Item Trials)		F-statistic	p	MSE	U (M)	C (M)
	Hint	3.37	.068	.042	.795	.746
	Comparison	1.21	.272	.042		
	Interaction	1.76	.186	.042		

ANOVA (One- and Two-Item Trials Difference Scores)						
	Hint	6.12	.014	.01		
	Comparison	.41	.523	.01		
	Interaction	.03	.86	.01		

Note: Conditions: C/U = Contrast/Unitary, C/C = Contrast/Compositional, M/U = Match/Unitary, M/C = Match/Compositional, U = Unitary Hint, C = Compositional Hint.

Taken together, these findings indicate that the unitary hint was just as effective when used with within- and between-category comparison (contrast condition), but the compositional hint was not, as using a compositional hint with between-category comparison is more effective than when it is used with within-category comparison (match condition). These results highlight the fact that the match/compositional condition was the least effective of the four conditions. Moreover, this finding goes against the prediction that a compositional hint would produce better learning under within- than between-category comparison. One possibility to consider is that within-category comparison is not particularly effective for relational concept discovery, but might be useful for relational concept recognition, particularly in cases where subjects already have a representation of the concept (e.g. recognizing that a novel concept is analogous to a familiar concept). Given that subjects in the present study initially did not know what defined the two categories, these subjects had to rely on the compositional hint in order to learn the category structures. However, this type of hint seems to require bottom-up learning and does not help to constrain (at least not initially) the type of interconnections among the ships that subjects consider, and thus might not be optimal to use in the match condition.

To test for whether the effect of hint depends on trial type, difference scores were computed between one- and two-item trials and a two-way ANOVA was conducted on these scores. In line with the hypothesized prediction, a statistically reliable interaction was found between type of hint and trial type, $F(1, 215) = 5.15$, $p = .024$, $MSE = .001$, as subjects who received a unitary hint outperformed subjects who received a compositional hint on one-item trials ($M_{unitary} = .82$; $M_{compositional} = .766$), but not on two-item trials ($M_{unitary} = .813$; $M_{compositional} = .79$). There was no interaction between trial type and comparison nor was there a three-way interaction between trial type, type of hint, and type of comparison, both $ps > .66$. These findings

support the idea that followed from Experiment 4, such that in cases where subjects have an additional learning aid that they can use in conjunction with a compositional hint, it can offset the benefit of using a unitary hint. In the present case, comparison seems to have indeed filled the role of said learning aid and helped subjects who received a compositional hint perform as well as those who received a unitary hint. Nevertheless, this benefit did not transfer to one-item trials, suggesting that these subjects may not have fully abstracted the relational structure of the two categories.

A.
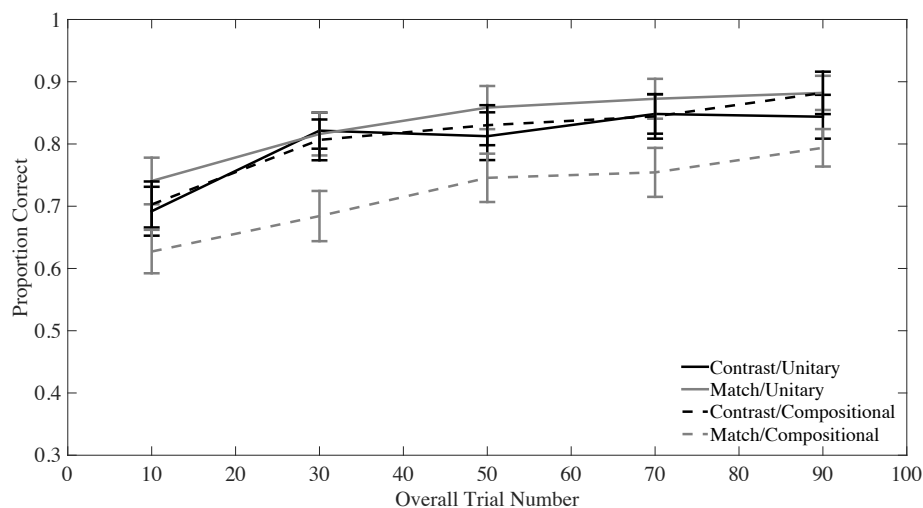
B.
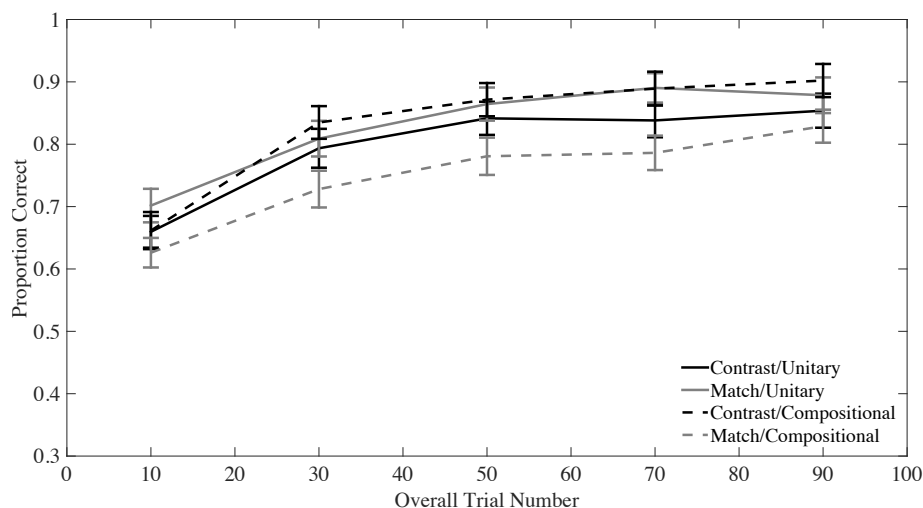
Figure 26. Average learning curves on one- and two-item trials and standard errors of the mean across blocks of 20 trials for each condition in Experiment 7. A. One-item trials. Each data point represents an average over two one-item trials. B. Two-item trials. Each data point represents an average over 16 two-item trials.

One possible explanation for these findings is therefore that the type of hint and comparison that subjects were provided in the other three conditions were more conducive to learning than in the match/compositional condition. It is important to note that the reason for this difference in learning may have varied among the three conditions that achieved better performance than subjects in the match/compositional condition. For instance, subjects in the unitary/match condition might have used the unitary hint as a top-down aid to construct a structured representation of the stimuli (familiarizing them with category structures), which might have then facilitated the alignment and abstraction of the shared structure between the two analogous items during within-category comparison. Thus, as in the unitary/inference condition in Experiment 2, these subjects might have been able to use the unitary hint to better represent the stimuli compositionally. In contrast, subjects in the contrast/unitary condition might have used the unitary hint to represent the categories unitarily. It has been proposed that such representations are represented similarly to features (Corral et al., 2017), and thus between-category comparison should highlight the diagnostic differences between the two categories, making subjects' classification judgments relatively straightforward. For these reasons, it is not surprising that subjects in these two conditions outperformed subjects in the match/compositional condition.

However, it is less clear as to why subjects in the contrast/compositional condition outperformed subjects in the match/compositional condition. For subjects who received a compositional hint, those in the match condition should have been able to align the corresponding elements between both scenarios and abstract their common structure, as follows from structure-mapping theory (Gentner, 1983). In contrast, full alignment was not possible for subjects in the contrast/compositional condition (because the two stimuli comprised different relational structures), thus these subjects should have performed worse than subjects in the match/compositional condition. One possibility raised by Corral et al. (2017) is that a failed alignment might highlight the relations that distinguish the two categories, and these relations can be represented as features. Consequently, subjects can then distinguish the two categories based on the presence or absence of those relations. For the present experiment, both categories consisted of the same relations, but differed in how those relations were linked. Subjects in the contrast condition might have therefore partially aligned the structures between the two categories (i.e., structural elements that were shared between the two categories), highlighting the differences between the two category structures and giving rise to corresponding unitary representations (e.g., Zorpes can be represented as an instance of shrinkage and Olatin can be represented as an instance of growth). This type of strategy thus allows for feature-based processing, which is posited to be less strenuous and more computationally efficient than using a compositionally structured representation (Forbus et al., 1995), which would require subjects to fully align the elements between the two stimuli (and represent their structure) on every trial. This latter process is posited to be computationally expensive and can place a large load on working memory (Kintsch & Bowles, 2002). Given these findings, it might therefore follow that discovering how the component parts of a stimulus are interrelated is easier when there are

structural differences between co-presented items than when those items have a common structure.

The present study used a classification task, but follow-up work should seek to replicate and extend these findings with an inference task. Experiment 2 showed that the benefit of the unitary hint is much stronger on inference than on classification, and thus the benefits of using a unitary hint might be more pronounced on an inference task. Moreover, it is possible that the present findings differ depending on the type of task that subjects complete (e.g., classification vs. inference). For instance, although the compositional hint was as effective for learning when used in conjunction with between-category comparison as the unitary hint, the unitary hint might prove to be more advantageous on inference-based learning. Future work will seek to address these questions.

Representational Preference and Individual Differences

Figure 27 shows the proportion of observed responses in each condition for each type of category rule. Unlike in the previous studies reported above, a binomial test did not reveal any differences in the proportion of observed responses between the unitary and compositional choices, $p = .685$. Similarly, a logistic regression did not find any differences in the type of category rule that subjects selected based on the type of comparison or hint that subjects used or an interaction between these two variables, all ps > .191. The type of category rule that subjects selected also did not predict performance on the classification task, $p = .640$ and no differences were found among conditions based on the type of category rules that subjects selected, all ps > .100.

Although no differences were found between the two types of category rules that subjects selected, these findings nevertheless help to support one of the ideas proposed in Experiment 1,

mainly that the type of descriptions that subjects select depend on the concept itself. In Experiments 1-4 and 6, subjects showed a strong preference for the descriptions of the concepts that conveyed unitized information. However, this preference was not found in the present study, which used different stimuli from those that were used in the aforementioned studies. One critical difference between the present stimuli and those used in the previous studies is that these stimuli were constructed in a way that subjects were required to examine and compare the relationships among each of the spaceships in each race to determine its category membership. Thus, each component in a stimulus (i.e., a single race) was itself made up of two components that shared a first-order relationship with one another (e.g., *beats*(*spaceship₁*, *spaceship₂*)). This additional layer of compositionality might have therefore encouraged compositional-based representations. Future work will need to be conducted to more directly test this idea.
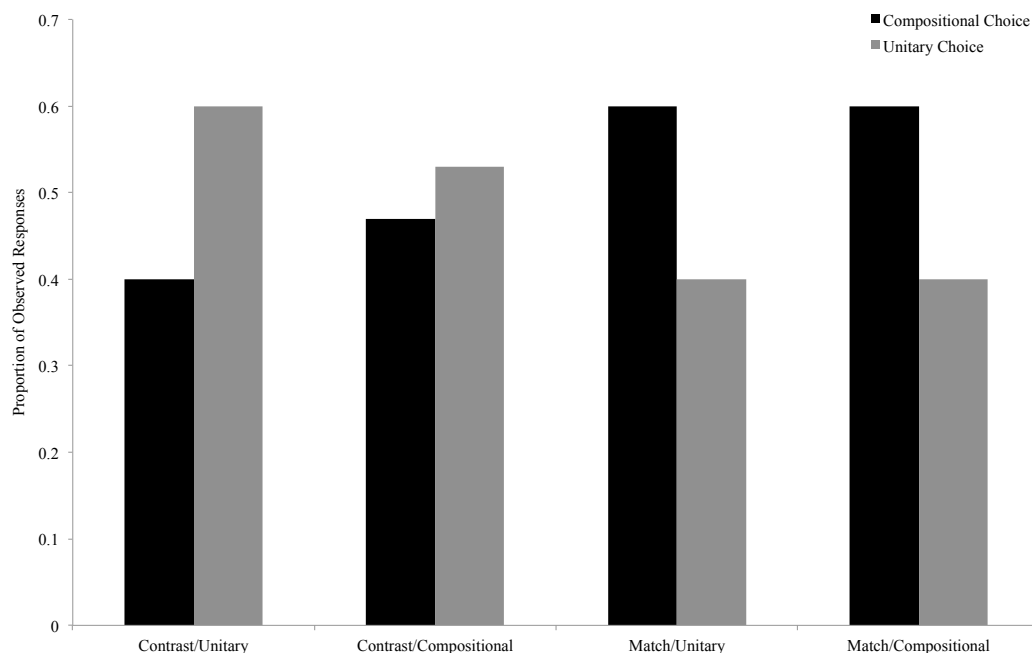
Figure 27. Shows the proportion of subjects who selected each the of category rules in each condition in Experiment 7.

CHAPTER IX

General Discussion

Although representation has been posited to play a critical role in relational learning (Chalmers et al., 1992; French, 1997, 2002; Markman, 1999; Markman & Dietrich, 2000; Markman & Gentner, 2000; Mitchell & Hofstadter, 1990), previous work has not directly examined this proposal. Nevertheless, for over 30 years empirical and theoretical work on relational learning has been premised on the assumption that people represent relational concepts compositionally (Doumas et al., 2008; Falkenhainer et al., 1989; Forbus et al., 1995; Gentner, 1983; Gick & Holyoak, 1983; Gentner & Markman, 1997; Holyoak & Thagard, 1997; Hummel & Holyoak, 1997, 2003; Markman, 1999; Markman & Gentner, 2000). The present paper puts this long held assumption to the test and reports seven experiments that investigate relational concept representation and how such representations affect relational learning. In accord with the main hypothesis advanced in this paper, the findings from these studies suggest that relational concepts can be represented in two fundamentally different ways, unitarily and compositionally. This possibility was acknowledged in early work on analogical reasoning (Gentner, 1983, Footnote 4), but its implications on relational concept acquisition have not been investigated until now.

Representational Preference

Experiments 1-4 and 6 provide evidence that contradicts the idea that people typically represent relational concepts compositionally, as subjects showed a strong preference for unitary-based descriptions of relational concepts over those that were described compositionally; no differences were observed in the types of category rules that subjects selected in Experiment 7. These studies used various types of relational concepts, which varied in structure, abstractness,

and familiarity. Experiment 1 presented subjects with 10 common relational nouns; Experiments 2-4 presented subjects with descriptions of the component parts of cleaning machines, which were defined by the interconnections (or lack thereof) among a machine's components; and Experiment 6 presented subjects with stimuli that were defined by how three arbitrarily chosen objects were related to one another based on size. Experiment 7 used categories that were defined by the relationships among three spaceships. Furthermore, these five studies used a variety of manipulations, including those that explicitly encouraged subjects to represent the stimuli compositionally (e.g., compositional hints in Experiments 2-4). Taken together, these studies suggest that people typically represent many relational concepts unitarily, and prefer this type of representation when both a unitary and compositional representation are available.

One objection that can be raised against this interpretation is that the unitary-based descriptions were simply written in a manner that was more accessible to subjects than the compositional descriptions, and thus the results do not reflect subjects' representational preference. However, Experiment 1 provides some evidence against this alternative account, as the types of definitions and scenarios that subjects provided predicted the type of definitions that they selected. Specifically, subjects who provided definitions and scenarios that referenced a greater number of the concepts' structural components selected a greater proportion of compositional-based definitions. If subjects' choices were based solely on differences in the writing quality of the two types of definitions, then their average compositionality scores should not have predicted the type of definitions that they selected. This relationship suggests that the description of the definitions and the responses that subjects generated both reflect the same underlying preference. Although it is unclear what this preference is, it seems reasonable to assume that this relationship reflects a representational preference. In sum, this finding suggests

that there are individual differences in how people represent relational concepts, which supports the primary hypothesis of this paper.

Putting aside the definitions that subjects selected in Experiment 1, the definitions that subjects generated perhaps provide the purest measure of how they represent the relational nouns that were used in this study (to the extent that these responses can provide such information). The results showed that these subjects generated definitions that were mostly unitary-based, providing further support that such concepts might be typically represented unitarily.

Moving on, as mentioned above, the findings from Experiment 7 showed that there were no differences in the type of category rules that subjects selected at the end of the study. This finding points to the idea that the type of representation that subjects prefer or can access for a relational concept can vary and might depend on the properties of the concept or stimulus. Experiment 1 used common relational nouns, which subjects were likely to be highly familiar with; Experiments 2-5 used more abstract stimuli that capitalized on subjects' prior knowledge about the secondary relations among a machine's component parts; and Experiment 6 used relational categories that instantiated the concepts *shrinkage* and *growth*. All of these studies therefore used concepts that subjects had some level of familiarity with. In such cases, subjects might be able to access and prefer to use unitary-based representations. In contrast, Experiment 7 used two categories, one that instantiated the concept of a *cycle* and the other of a *hierarchy*. Although these concepts might be familiar or readily accessible to some subjects, they might be less familiar or more difficult to access for others, which might have made it less likely for these latter subjects to represent the categories unitarily. This possibility might explain the reason that there were no differences in the types of category rules that subjects selected in Experiment 7.

Thus, it is possible that subjects prefer compositional-based representations in cases where they are unfamiliar with a given concept.

Taken as a whole, the findings reviewed in this subsection support the idea that relational concepts can indeed be represented in two fundamentally different ways, unitarily and compositionally, and suggest that people might typically represent such concepts unitarily. Nevertheless, it is acknowledged that the responses provided by subjects (the scenarios and definitions that were selected and generated by subjects in Experiment 1 and the category rules that were selected in Experiments 2-4, 6, and 7) might reflect something other than how they represent the corresponding concepts. Thus, although the present findings can be said to be fairly reliable and somewhat suggestive, they are far from conclusive and more work will be required in order to draw a more definitive conclusion.

Representation and Relational Concept Learning

Experiment 2 showed that both unitary- and compositional-based hints could be used to aid relational learning (as compared to control groups), otherwise no differences in performance would have been observed. However, the benefit of the compositional hint seems to be restricted to classification learning, as it was seemingly ineffective on the inference task. Moreover, the unitary hint led to better learning on both classification and inference than the compositional hint, suggesting that unitary-based representations might be more conducive to learning. Experiment 3 replicated these findings using a classification task and also showed that as long as subjects are provided a unitary hint, even if they are shown a compositional hint for most of the study (as subjects in the unitary/switch condition were), they will perform better than subjects who do not receive a unitary hint at all. The benefit of the unitary hint was also observed in Experiment 4, in which subjects who switched from a compositional to a unitary hint showed

rapid improvement in learning when they switched from classification to inference and were the only group of the four (compositional/switch, compositional/no-switch, and unitary/no-switch) that showed a statistically reliable improvement in performance when they switched over to the other task.

The ineffectiveness of the compositional hint in the inference task in Experiment 2 might be thought of as somewhat surprising, given that relational inference learning requires subjects to explicitly consider how the component parts within a given stimulus are interconnected. However, it is possible that because compositional processing is working memory intensive (Forbus et al., 1995; Kintsch & Bowles, 2002), a compositional hint is not particularly useful in cases where people must explicitly discover novel connections among relations. In such cases, a unitary representation (if it is available) might be particularly helpful, as it might constrain the types of relations that are considered and provide structure to how such relations are interconnected. The unitary hint might derive its effectiveness from its ability to leverage subjects' prior knowledge, which they can then apply in novel ways to discover the concepts that define the experiment's categories. In contrast, a compositional hint might not provide this type of conceptual guidance.

Interestingly, no differences in performance were observed among any of the groups in the inference task in Experiment 4. Indeed, subjects in the compositional/no-switch condition performed as well as subjects in the other three groups. This finding is meaningful because in Experiment 2 there were large differences in inference performance between subjects who received a unitary hint and subjects who received a compositional hint. One possible reason as to why a similar result was not observed in Experiment 4 is that subjects engaged in classification before inference, which might have helped them to learn the category structure of a Morkel (or at

least part of it). Consequently, subjects might have been able to use this knowledge in a top-down manner during the inference task, offloading some of the processing strain from engaging in inference and representing the category compositionally. These findings suggest that once a conceptual aid has been acquired, a compositional representation can be just as effective as a unitary concept. This idea is supported by the findings in Experiment 7, which replicated and extended the findings from Experiment 2, showing that the a unitary hint led to better learning than the compositional hint on one-item trials, but not on two-item trials. Because two-item trials allow for comparison, they might have filled a similar role as classification did in Experiment 4, in that both provide a learning aid that can reduce the computational cost of using a compositional representation.

This idea might also account for the reason that subjects in the match/compositional condition in Experiment 7 performed worse than subjects in the other three conditions, as structural alignment might place a large strain on working memory, and thus matching the corresponding elements between two analogous items (as both the compositional hint and the match condition seem to encourage) and abstracting their common structure might be particularly challenging (as discussed above). As a result, this condition might lack a buffer against the processing strain that comes from structural alignment and representing the categories compositionally, whereas in the other three conditions subjects were either provided a unitary hint or might have been able to learn through failures of alignment (compositional/contrast condition), either of which might lead to more efficient processing (as described above).

One critique that can be raised against the hints that were used in the studies presented here is that the compositional hint instructed subjects to look for relations without providing them any information about what the correct relations were, whereas the unitary hint instructed

subjects to look for a particular type of global property. However, such a critique ignores that the unitary hint that subjects were provided was much more ambiguous than the compositional hint, as a machine that is built *intuitively* (Experiments 2-4) or an outcome of a tournament that *follows logically* (Experiment 7) is a vague and unspecified set of instructions, as that which is intuitive or follows logically can be highly subjective and can widely vary from one subject to another. Moreover, just about any stimulus or concept can be made to fit these properties (including non-Morkels). Thus, the constraint and structure that are provided from a unitary hint can be said to be implicit, in that the hint is devoid of explicit structure and there is no explicit information that is specific to the stimuli (because the hint can more or less apply to anything). In contrast, the compositional hint is much more specific and not only tells subjects to look for relationships, but tells them what components to look for relationships among.

Furthermore, for both a unitary and a compositional hint, subjects are required to figure out how the hint applies to a stimulus. Specifically, subjects who received a unitary hint were required to figure out how or why a machine is intuitive or would function, which was based on how the machine's component parts were interconnected. Likewise, subjects who received a compositional hint were required to figure out what the relationships were among the machine's components. Thus, in both cases subjects were required to figure out the same information, which was how the machine's component parts were interconnected. This point holds particularly true for the inference condition in which subjects had to explicitly infer the machine's missing component, which was based on its relationship to the other two components. Moreover, the unitary hint was only specific to the Morkel items and produced no useful information about non-Morkel items, whereas the compositional hint should have been useful for both types of stimuli. Specifically, examining the relationships among the component parts of a

stimulus should have equally benefitted the learning of Morkel and non-Morkel items. Nevertheless, subjects who received the unitary hint outperformed subjects on non-Morkel inference items, despite receiving less explicitly applicable information. These findings argue against the idea that the unitary hint is more explicitly informative than the compositional hint, and seems to provide evidence to the contrary.

Summary: Experiments 1-7

Experiment 1 examined whether three different types of manipulations, which were intended to encourage subjects to represent relational nouns compositionally, could affect the types of definitions that subjects selected for each noun; these conditions were compared to a control condition. In line with one of the primary predictions, subjects in the control condition showed a strong preference for the unitary definitions. However, none of the manipulations seemed to affect the types of definitions that subjects selected, as subjects in these three conditions showed a strong preference for definitions that were described unitarily. This outcome goes against the prediction that the manipulations used in this study would change how subjects represented the relational nouns, suggesting that such changes in representation might be fairly challenging and likely requires stronger manipulations.

Experiments 2-7 examined the relationship between representation and relational learning. Specifically, Experiment 2 tested the congruence between type of hint and task type and found an interaction, such that in general, a unitary hint leads to better learning than a compositional hint, but this benefit is particularly strong on an inference task. The finding that a unitary hint produced better classification learning than the compositional hint (as shown in Experiments 2-3) is in line with the a priori prediction that subjects should be able to recognize a unitary attribute directly in a stimulus, similarly to a feature, which should be less

computationally taxing than learning the categories based on their relational structure. However, the finding that the unitary hint led to better inference learning than the compositional hint was unexpected and contradicts the prediction that because relational inference requires understanding the relationships among a stimulus's component parts, a compositional hint should lead to better learning than a unitary hint. It is speculated that subjects who received a unitary hint were able to use that hint in a top-down manner to build up a structured representation, which is less computationally expensive than discovering the structured concept in a bottom-up manner, as subjects in the compositional hint were required to do. This post hoc hypothesis led to specific predictions that were more directly tested in Experiment 7 (as discussed below).

Experiment 3 had two specific predictions, such that subjects in the unitary/no-switch condition and subjects in the compositional/switch condition would outperform subjects in the compositional/no-switch condition. These predictions were premised on the idea that the unitary hint should lead to superior category learning, as discussed above and as shown in Experiment 2. Because subjects in the unitary/no-switch and the compositional/switch conditions were either exclusively or primarily shown the unitary hint, they were expected to outperform subjects in the compositional/no-switch condition, in which only a compositional hint was used. These predictions were indeed supported by the data, and taken with the findings from Experiment 2, provide further support for the idea that unitary representations lead to superior category learning than compositional representations. Although there were no a priori predictions about the unitary/switch condition, subjects in this condition marginally outperformed subjects in the compositional/no-switch condition, suggesting that when subjects have access to both types of representations they might abandon the compositional representation in favor of the unitary representation.

Additionally, based on the findings from Experiment 2, it was predicted that subjects in the unitary/no-switch condition would outperform subjects in the compositional/no-switch condition on the inference task in Experiment 4. However, this hypothesis was not supported as no differences in performance were found between subjects in these two conditions. It is speculated that the classification task that subjects engaged in before completing the inference task may have served as a learning aid for subjects who received the compositional hint, which allowed these subjects to learn the categories just as well as subjects who received the unitary hint. The hypothesis that using a compositional hint in conjunction with a learning aid can offset the benefits of presenting subjects a unitary hint was directly tested in Experiment 7.

Experiments 5 and 6 examined the congruence between type of comparison and task type. As in Experiment 2, Experiments 5a and 5b revealed an interaction, this time between comparison and task type, such that within-category comparison led to better learning than between-category comparison on an inference task, whereas between-category comparison led to slightly better learning on a classification task. These findings are exactly in line with the predictions outlined in Experiment 5a. The match advantage provides support for the idea that inference-based learning encourages compositional representations and that subjects are able to fully align and abstract the common structure between two items, as predicted by structure-mapping theory (Gentner, 1983; Lassaline, 1996; Lassaline & Murphy, 1998). The slight contrast advantage provides mild support for the idea that classification encourages unitary-based representations, which can be treated similarly to features, and the contrast highlights the discriminative unitary attributes between the two categories, which are less computationally expensive than compositional representations and structural alignment. In sum, the interaction found in Experiments 5a and 5b suggest that the two types of tasks encourage different types of

representations, which can be leveraged differently during relational learning, depending on the types of comparisons that subjects engage in.

Experiment 6 used stimuli that were designed with the intention of inhibiting a unitary representation from emerging, but did not yield statistically reliable results. Nevertheless, this study did show a trend in the direction of a match advantage (which was expected if subjects did not have access to a unitary representation) for both task types (and no interaction), which might suggest the interaction observed in Experiments 5a and 5b might itself depend on the properties of the stimuli. Nevertheless, the findings from Experiment 6 did not provide support for the a priori prediction of a match advantage on both the classification and inference tasks and further work will be required to better understand this inconclusive finding.

Lastly, Experiment 7 used a classification task and examined the congruence between type of comparison and type of hint, and once again revealed an interaction between these two factors, suggesting that the type of hint that is most effective for classification learning depends on the type of comparison that subjects engage in (as discussed above). Although this interaction was expected, some of the group differences were not. The prediction that subjects who received a unitary hint would perform better in the contrast condition than in the match condition was not supported, as no differences in performance were found between these two groups. Moreover, evidence was found against the hypothesis that subjects who received a compositional hint would perform better in the match condition than in the contrast condition, as the opposite pattern was found, providing support for the prediction that arose from the alignable differences hypothesis advanced by Corral et al. (2017). Subjects in the unitary/match condition also outperformed subjects in the compositional/match condition. This finding is directly in line with the a priori prediction that was based on the findings from Experiment 2 (in which the unitary

hint led to better inference learning than the compositional hint), from which it would be expected that a unitary-based representation can be used to construct a structured concept in a way that is more computationally efficient than building a compositional representation in a bottom-up manner. Thus, subjects in the unitary/match condition should be expected to perform better than subjects in the compositional/match condition. Because subjects in the match conditions are posited to learn through structural alignment, subjects who received a unitary hint might have therefore been better able to align and abstract the common structure between two items than subjects in the compositional/match condition. Additionally, subjects in the unitary/contrast condition outperformed subjects in the compositional/match condition. Although this outcome was not explicitly predicted, it is in line with the idea that a unitary concept can be readily recognized from between-category comparison, which should lead to more efficient processing than representing a concept compositionally and engaging in structural alignment. On a last note, Experiment 7 also provided support for the prediction that followed directly from Experiment 4, which was that the benefit of the unitary hint over the compositional hint can be offset when the compositional hint is used in conjunction with a learning aid (e.g., comparison), as subjects who received a unitary hint outperformed subjects who received a compositional hint on one-item trials, but not two-item trials.

In sum, although none of these studies by themselves provided conclusive evidence for a specific hypothesis, together they do provide converging evidence for the primary hypothesis advanced in this paper. Mainly, these studies support the idea that relational concepts are represented in two fundamentally different ways, and the type of representation that leads to better relational learning depends on various factors, such as task type and comparison. Furthermore, these reported interactions paint a complicated picture going forward, which is

further complicated by the findings from Experiment 1, which indicate that there are likely individual differences in how subjects represent relational concepts. Additionally, the results from Experiment 6 point to the possibility that these interactions might themselves depend on the type of concept that subjects are learning. Hence, the types of interventions that are most effective for improving relational learning might vary by instruction (e.g., type of hint), task type, type of comparison, concept type, and individuals. This take away is not particularly surprising given the complicated nature of representation and relational learning. Nevertheless, future work on relational learning will need to take these issues into careful consideration when designing experiments.

Novel Concept Learning Versus Concept Recognition

One important issue that has yet to be addressed is that there is a distinction between learning a novel concept and learning which concepts (among those the subject has a priori knowledge about) define a given category.[13] For instance, it can be argued that the latter type of learning occurred in Experiments 2-5, because subjects presumably had pre-existing knowledge about the secondary relations among the Morkel machines' component parts. A similar point can be made about Experiments 6 and 7, as subjects likely had knowledge about the concepts of

---

[13] It is important to acknowledge that the conditions under which novel concept acquisition truly occur are vague and underspecified, and depend on how novel learning is defined (for which there are many reasonable definitions). This question has long been debated in philosophy and has been a fairly contentious topic, as many prominent scholars (e.g., Plato, Descartes) have argued that novel learning never actually occurs (Gorham, 2002; Stich, 1975). These issues are not given further consideration here, as they are beyond the scope of the present paper.

shrinkage and *growth* (Experiment 6) and the concepts of *cycle* and *hierarchy* (Experiment 7). It is important to note that structure-mapping theory (Gentner, 1983) is primarily applicable to these latter cases, as the theory holds that the elements between the base (a concept that the subject already has in memory) and target (a novel concept or scenario) are aligned and their common structure is abstracted. This process allows for inference projection from the base to the target (Spellman & Holyoak, 1992, 1996), such that knowledge from a previous concept or scenario is applied to a novel situation. Indeed, previous work has raised this exact point and argued that structure-mapping theory, on its own, does not account for how relational concepts are initially acquired (Chalmers et al., 1992; French, 1997; Mitchell & Hofstadter, 1990).

Moreover, even in cases where subjects are asked to learn a category that instantiates a concept that is known, it is erroneous to conclude that novel relational learning is not taking place. Extensive work has shown that people often fail to recognize when a presumably known relational concept is presented in a novel context (Gick & Holyoak, 1983; Gentner & Schumacher, 1986; Holyoak & Koh, 1987; Reed, 1989; Ross, 1987, 1989). In such cases, learning can only proceed by the subject discovering the concept anew and should consequently be driven by the same type of learning mechanisms that underlie novel relational concept acquisition. For these reasons, although Experiments 2-7 used concepts that subjects likely had some familiarity with, it is likely that many (if not most) subjects had to learn these concepts from scratch, as they were likely unaware (at least until after the concept was learned) that the concepts they were learning were instantiations of the concepts they had some prior knowledge about. It also seems somewhat unlikely that subjects had prior knowledge about the non-Morkel items in Experiments 2-5 (given that the relational rule that defined these concepts were made up by the researcher), which subjects were explicitly required to discover in the inference task.

Thus, it seems a fair to conclude that the present findings are likely applicable to both novel concept acquisition and learning which concepts from memory define a given category (i.e., analogical learning).

Relational Reasoning and Concept Representation in the Brain

Although the studies presented here provide support for the idea that relational concepts can indeed be represented in two fundamentally different ways, and that people might typically prefer to represent such concepts unitarily, there is one potential shortcoming to these studies that requires consideration. Although subjects were encouraged to represent the stimuli unitarily or compositionally, it cannot be known for certain whether subjects adopted either of these representations. This issue has historically plagued cognitive scientists and highlights the need for improved assessment on concept representation. This challenge might account for the reason that researchers have typically neglected the role that representation plays in relational and analogical learning (Chalmers et al., 1992; French, 1997; Mitchell & Hofstadter, 1990), despite recognizing its importance (Markman, 1999; Markman & Gentner, 2000). Although there is not a perfect solution to this problem, the technological advances over the last two decades in imaging techniques make the field of cognitive neuroscience relatively well suited for addressing this issue. Specifically, if relational concepts can indeed be represented unitarily and compositionally, one would expect these two types of representations to yield different types of neural patterns of activation.

Neural Correlates of Relational and Analogical Reasoning

Although not particularly plentiful, there are studies that have examined the neural correlates of relational and analogical reasoning (e.g., Bunge, Wendelken, Badre, & Wagner,

2005; Green, Fugelsang, Kraemer, Shamosh, & Dunbar, 2006; Green, Kraemer, Fugelsang, Gray & Dunbar, 2010; Mason & Just, 2016). A relatively consistent finding is that the prefrontal cortex (PFC) plays a prominent role in relational reasoning. For instance, various studies have shown high levels of activation in the PFC on a multitude of relational reasoning tasks (e.g., solving problems in mechanical physics, phase changes of matter, relationship among electrical circuits; Dunbar, Fugelsang, & Stein, 2007; Brault Foisy, Potvin, Riopel, & Masson, 2015; Masson, Potvin, Riopel, & Brault Foisy, 2014; Masson, Potvin, Riopel, Brault Foisy, & Lafortune, 2012). Conceptually related work has shown a strong relationship between activation in the PFC and relational concept acquisition (concepts about air-pressure) and scientific reasoning (Kwon & Lawson, 2000). Other work has compared the neural activation of experts and novices during various types of relational reasoning tasks and has shown that experts have greater activation in the PFC than novices during problem solving within the experts' domain of the expertise (Amalric & Dehaene, 2016; Foisy et al., 2015; Nelson, Lizcano, Atkins, & Dunbar, 2007). Another study found that subjects who were asked to determine whether geometry problems (some of which were analogous) were similar to one another had greater activation in the left dorsomedial PFC than when they were asked to make literal comparisons (e.g., comparing whether two statements were identical; Wharton et al., 2000).

These findings are in accord with previous work, which suggests that the PFC plays an important role in the use of rules (Miller, Nieder, Freedman, & Wallis, 2003; Schoenbaum & Setlow, 2001; White & Wise, 1999) and the integration of relational information (Christoff et al., 2001; Morrison et al., 2004), both of which are critical for relational reasoning. Integrating the way in which multiple relations and objects are bound together into a single representation arguably lies at the heart of relational reasoning (Holyoak & Thagard, 1997). Furthermore, these

relationships often follow a given set of rules (e.g., causal structures) or require the learner to apply a given set of rules to solve a particular scenario (e.g., applying a solution strategy to a given problem type).

Imaging research on analogical reasoning has provided somewhat more detailed findings about which parts of the PFC are involved in relational reasoning. Specifically, activation in the left frontopolar cortex (FPC) has been shown to increase during analogical reasoning tasks in comparison to control conditions (working memory tasks that do not require analogical reasoning; Green et al., 2006). Related work has shown a positive relationship between activity in the lateral FPC and the number of relations that subjects are asked to process (Cho et al., 2010). That is, as the number of relations for a given analogy between two items increases, so does activity in the lateral FPC. Other studies that have manipulated the number of relations for a given analogy have shown similar findings in respect to the increased activation in the left lateral FPC (Christoff et al., 2001; Kroger et al., 2002). These results have been interpreted as providing support for the idea that the left FPC is primarily responsible for integrating relational information (Ramnani & Owen, 2004).

The left FPC has also been implicated as playing a central role in mapping and aligning the corresponding elements between two analogous concepts or scenarios (Green et al., 2006). Research has shown that when the semantic distance among the relations between two analogous concepts is increased (i.e., going from a within- to a between-domain analogy), so too does activity in the left FPC (Green et al., 2010; Green, Kraemer, Fugelsang, Gray & Dunbar, 2012; Krawczyk, McClelland, Donovan, Maguire, & Tillman, 2010; Wendelken, Nakhabenko, Donohue, Carter, & Bunge, 2008). Much of this work has also shown that activation in the left

FPC is stronger in the anterior regions when the number of relations or semantic distance between analogous concepts increases.

Taken together, these findings suggest that the left FPC plays a prominent role in representing relational concepts compositionally, as activation in this region seems to increase with the demand for representing the structural components of a given relational concept. This idea is further supported by findings which have shown that there is an increase in the activation of the left FPC in cases where subjects are asked to remember specific components about an episodic event (Dobbins, Foley, Schacter, & Wagner, 2002; Nolde, Johnson, & D'Esposito, 1998; Ranganath, Johnson, & D'Esposito, 2000). This finding is particularly relevant because episodic memories are structured, such that each object within a given episode is bound to a specific role, which they fill at a specific point within the given episode. Thus, in cases where a subject represents a given stimulus compositionally, activation in the left FPC should be expected to increase in comparison to cases where the subject represents the concept unitarily. Furthermore, the FPC has also been found to play an important role in inference-based reasoning and in cases where subjects are required to generate information (Christoff & Gabrieli, 2002). This finding suggests that subjects who are asked to engage in inference-based reasoning or who use a unitary hint to generate the structure of a stimulus (as it was proposed that subjects in Experiments 2 and 7 might have done) might be expected to have greater activation in the FPC than subjects who are either engaged in classification or who do not use this strategy.

This review raises the question of which brain regions are involved when a relational concept is represented unitarily. An answer to this question might be found from a study conducted Bunge et al. (2005). On some trials, subjects were asked to determine whether two word pairs (e.g., bouquet/flower and chain/link) were analogous or whether two words were

related (rain and draught). As in the findings reviewed above, Bunge et al. report an increase in activation in the left FPC on trials where subjects were required to make an analogy judgment, suggesting that subjects may have been representing such stimuli compositionally. In contrast, on trials in which subjects made similarity judgments between two words, there was an increase in activation in the anterior left inferior prefrontal cortex (aLIPC). For these trials, subjects could simply retrieve a shared relation between the two words. For instance, if a subject is asked to make a relatedness judgment between chain and link, a correct judgment simply requires the subject to retrieve the *contain* relation (because a chain contains a link). Related work has shown the aLIPC is important in the retrieval of semantic knowledge, including relations (Jackson, Hoffman, Pobric, & Lambon Ralph, 2015; Noonan, Jefferies, Visser, & Lambon Ralph, 2013; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner, Maril, Bjork, & Schacter, 2001). Bunge et al. propose that the aLIPC plays a critical role in retrieving common relations between analogous concepts. Because relations might be represented unitarily, similarly to a feature (Corral et al., 2017), representing a relational concept unitarily might produce greater levels of activation in the aLIPC than when the concept is represented compositionally.

It is important to point out that the analogy task used by Bunge et al. (2005) could potentially be completed without structural alignment, such that a subject could simply recognize that the word pairs share the same relation (e.g., bouquet *contains* flowers and chain *contains* link). Solving the task in this manner would require the retrieval of common relations and thus should produce an increase in the activation of the aLIPC. Although Bunge et al. report an increase in activation in the left FPC during their analogical reasoning task, this increase might simply reflect that there were more components to process in the analogy task (four words) than in the word similarity task (two words), as activation in this region seems to increase as the

number of components that must be processed for a given task increases. It is therefore not clear whether the increase in activation in the left FPC during the analogy task reported by Bunge et al. were due to subjects engaging in structural alignment (as proposed by Bunge et al.) or due to a general increase in the processing of a stimulus' component parts. Imaging work on analogical and relational reasoning has not considered the possibility that many of these tasks can be completed using unitary-based representations, as this work has been primarily premised on a structure-mapping account of analogical reasoning (Gentner, 1983). Further work is therefore necessary to provide further clarity on this issue.

This section reviews imaging research on relational and analogical reasoning. From this review there are specific predictions that follow about which brain regions are likely to show greater activation if a subject represents a given concept unitarily or compositionally. Although these predictions have yet to be directly tested, imaging research has the potential to address what is possibly the most daunting challenge in research on concept representation, which involves accurately assessing how subjects represent a given concept. To briefly summarize the findings reviewed in this section, a strong relationship has been found between activity in the left FPC and analogical reasoning, as well as relational retrieval and activity in the aLIPC. These findings lead to two predictions. One is that there should be a greater amount of activation in the left FPC when subjects represent a given concept compositionally. The other prediction is that there should be stronger activation in the aLIPC when they represent a concept unitarily. It is important to note that these are the exact pattern of results that were reported by Bunge et al. (2005) in the analogy and semantic similarity conditions, respectively, the former of which might have been conducive to representing the stimuli compositionally and the latter to representing the stimuli unitarily. Nevertheless, directly testing these predictions in an imaging study is perhaps a

necessary step in the path to providing converging evidence for the primary idea advanced in this

paper, which is that relational concepts can be represented in two fundamentally different ways.

CHAPTER X

Conclusion

This paper reports seven studies that provide support for the idea that relational concepts can be represented in two different ways, compositionally and unitarily, and moreover that various factors might affect how a concept is learned and represented. Although the findings reported above are suggestive, the results are far from conclusive (and can be interpreted in numerous ways). An extensive amount of work will therefore be required to answer the questions there were raised here. Nevertheless, this research has laid the groundwork for such follow-up work to be conducted.

Furthermore, the findings reported here (and any subsequent related research) seem particularly applicable to education and instruction, as they might provide insight into how different types of descriptions for a given relational concept can affect students' representations, as well as how such representations affect learning. Indeed, students are often required to learn various types of structured concepts, and must often engage in both classification and inference. For instance, in mathematics, students must recognize various instantiations of a given problem type, a process that relies on classification, and must also make inferences about how to apply a given solution. These findings thus hold the potential to improve how relational concepts are taught in the classroom.

Furthermore, the present findings have theoretical implications for relational concept learning and representation, and have the potential to affect current theories of analogical reasoning and learning. In particular, research within the theoretical framework of structure mapping (Doumas et al., 2008; Hummel & Holyoak, 2003) has placed a heavy emphasis on alignment processes operating on compositional representations, but our findings suggest that

subjects more naturally represent such concepts unitarily, and that such representations produce a greater and more robust benefit to learning. During comparison of two scenarios, if the critical information can be represented unitarily, then there is no need for structural alignment, because the two can be recognized through the same sort of processing that is possible with feature-based representations, that is, flat (setwise) comparison to identify which properties they have in common. To be clear, this proposal is not intended to argue against the idea that structural alignment of compositional representations plays a prominent role in the more impressive feats of human reasoning (e.g., creativity or scientific discovery), but rather to point out that in more mundane cases, simpler processes and representations may be involved. Nevertheless, further work is necessary to better understand which conditions facilitate unitary and compositional representations.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*, 87-113.

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians, *Proceedings of the National Academy of Sciences*, *113*, 4909-4917.

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews: Neuroscience, 4,* 829-839.

Boysen, S. T., Bernston, G., Hannan, M. & Cacioppo, J. (1996) Quantity-based inference and symbolic representation in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 76–86.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61-100.

Bunge, S.A., Wendelken, C., Badre, D., & Wagner, A.D. (2005). Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cerebral Cortex*, *15*, 239-249.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306–307.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental and Theoretical and Artificial Intelligence, 4*, 185-211.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55-81.

Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121–152.

Cho, S., Moody, T., Fernandino, L., Mumford, J., Poldrack, R., Cannon, T., Knowlton, B., & Holyoak, K. (2010). Common and dissociable prefrontal Loci associated with component mechanisms of analogical reasoning. *Cereb Cortex, 20*, 524-533.

Christoff, K., & Gabrieli, J. D. E. (2002). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology*, *28*, 168-186.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli, J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, *14*, 1136–1149.

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, *10*, 370–74.

Clement, C. A., & Gentner, D. (1991). Systematicity as a selection constraint in analogical mapping. *Cognitive Science, 15*, 89-132.

Corral, D., & Jones, M (2014). The effects of relational structure on analogical learning. *Cognition, 132*, 280-300.

Corral, D., & Jones, M. *Carving analogy by its joints*. In preparation.

Corral, D., Kurtz, K. J. & Jones, M. (2017). *Learning relational concepts from within- vs. between-category comparisons*. Under revision.

Czerwinski, M. P., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology: Special Issue on Views and Varieties of Automaticity, 105*, 271–315.

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

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: multiple prefrontal processes subserve source memory. *Neuron, 35*, 989–996.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1-43.

Dunbar, K. N., Fugelsang, J. A., & Stein, C. (2007). Do naïve theories ever go away? Using brain and behavior to understand changes in concepts. In M. C. Lovett & P. Shah (Eds.), *Thinking with Data: 33rd Carnegie Symposium on Cognition* (pp. 193-206). Mahwah, NJ: Erlbaum.

Ericsson, K. A., Chase, W. G. & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181–82.

Ericsson, K. A., Krampe, R. Th., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Falkenhainer, B., Forbus, K. & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence, 20*, 1–63.

Brault Foisy, L.-M., Potvin, P., Riopel, M., & Masson, S. (2015). Is inhibition involved in overcoming a common physics misconception in mechanics? *Trends in Neuroscience and Education*, *4*, 26–36.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141-205.

Forbus, K. D., Gentner, D., Markman, A. B., & Ferguson, R. W. (1998). Analogy just looks like high-level perception: Why a domain-general approach to analogical mapping is right. *Journal of Experimental and Theoretical Artificial Intelligence, 10*, 231-257.

Foster, J. M., Cañas, F., & Jones, M. (2012). Learning conceptual hierarchies by iterated relational consolidation. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*.

French, R. M. (1997). When coffee cups are like old elephants or why representation modules don't make sense. In A. Riegler & M. Peschl (Eds.), *Proceedings of the International Conference on New Trends in Cognitive Science* (pp. 158-163). Austrian Society for Cognitive Science.

French, R. M. (2002). The Computational modeling of analogy-making. *Trends in Cognitive Science, 6*, 200-205.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gentner, D., & Bowdle, B. (2008). Metaphor as structure-mapping. In R. Gibbs (Ed.) *The Cambridge Handbook of Metaphor and Thought* (pp. 109-128). New York, NY: Cambridge University Press.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*, 266-276.

Gentner, D., & Kurtz, K. (2006). Relations, objects, and the composition of analogies. *Cognitive Science*, *30*, 609-642.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *5*, 152-158.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist, 52*, 45-56.

Gentner, D., & Markman, A. B. (2005). Defining structural similarity. *Journal of Cognitive Science*, *6*, 1–20.

Gentner, D., & Schumacher, R. M. (1986). Use of structure mapping theory for complex systems. *Proceedings of the 1986 IEEE International Conference on Systems, Man, and Cybernetics*, 252-258.

Gentner, D., & Wolff, P. (1997). Alignment in the processing of metaphor. *Journal of Memory and Language*, *37*, 331-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Gildea, P., & Glucksberg, S. (1983). On understanding metaphor. The role of context. *Journal of Verbal Learning and Verbal Behavior, 22,* 577-590.

Giora, R. (1999). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics, 31*, 919–929.

Glucksberg, S. (2003). The psycholinguistics of metaphor. *Trends in Cognitive Sciences, 7*, 92–96.

Glucksberg, S., Gildea, P., & Bookin, H. (1982). On understanding nonliteral speech: Can people ignore metaphors? *Journal of Verbal Learning and Verbal Behavior, 21,* 85-98.

Glucksberg, S., & Keysar, B. (1990). Understanding metaphorical comparisons: Beyond similarity. *Psychological Review, 97*, 3-18.

Glucksberg, S., McGlone, M. S., & Manfredi, D. (1997). Property attribution in metaphor comprehension. *Journal of Memory and Language*, *36*, 50-67.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236-243.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, pp. 86–112.

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 29-50.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology, 23*, 222-264.

Gorham, G. (2002). Descartes on the innateness of all ideas. *Canadian Journal of Philosophy, 32*, 355-388.

Green, A.E., Fugelsang, J.A., Kraemer, D.J., Shamosh, N.A., & Dunbar, K.N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Research*, *1096*, 125–137.

Green, A.E., Kraemer, D.J., Fugelsang, J.A., Gray, J.R., & Dunbar, K.N. (2010). Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cerebral Cortex, 20*, 70–76.

Green, A.E., Kraemer, D.J.M., Fugelsang, J.A., Gray, J.R., & Dunbar, K.N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 264-272.

Harris, R. J. (1976). Comprehension of metaphors: A test of the two-stage processing model. *Bulletin of the Psychonomic Society*, *8*, 312-314.

Higgins, E. J. (2012). Comparing comparisons in category learning. *Doctoral Dissertation, University of Illinois*.

Healy, A. F. (1976). Detection errors on the word *the*: Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception & Performance*, *2*, 235–242.

Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332-340.

Holyoak, K. J., & Thagard, P. (1997). The analogical mind. *American Psychologist, 52*, 35-44.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427-466.

Hummel J. E., & Holyoak K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220-264.

Jackson, R. L., Hoffman, P., Pobric, G., & Lambon Ralph, M. A. (2015). The nature and neural correlates of semantic association vs. conceptual similarity. *Cerebral Cortex, 25*, 4319–4333.

Keane, M. T. (1997). What makes an analogy difficult? The effects of order and causal structure on analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 946-967.

Keil, F. C. (2003a). Folkscience: Coarse interpretations of a complex reality. *Trends in Cognitive Science*, *7*, 368-373.

Keil, F. C. (2003b). Categorization, causation, and the limits of understanding. *Language and Cognitive Processes, 18*, 663-692.

Keil, F. C. (2006). Explanation and understanding. *Annual Review of Psychology*, *57*, 227-254

Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol*, *17*, 249–262.

Kokinov, B. (1990). Associative memory-based reasoning: Some experimental results. *Proceedings of the 12th Annual Conference of the Cognitive Science Society*. Erlbaum, Hillsdale, NJ.

Kokinov, B. (1994). Flexibility versus efficiency: The dual answer. In P. Jorrand & V. Sgurev (Eds.), *Artificial intelligence: Methodology, systems, applications*. Singapore: World Scientific Publ.

Kokinov, B., & French, R. (2003). Computational Models of Analogy-Making. In: Lynn Nadel (ed.) Encyclopedia of Cognitive Science. Macmillan.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development*, *67*, 2797-2822.

Krawczyk, D. C., McClelland, M. M., Donovan, C. E., Maguire, M. J., & Tillman, G. D. (2010). An fMRI investigation of cognitive stages in reasoning by analogy. *Brain Research, 1342*, 63-73

Kroger, J. K., Sabb, F. W., Fales, C. L., Bookheimer, S. Y., Cohen, M. S., Holyoak, K. J. (2002). Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. *Cerebral Cortex, 12*, 477-485.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22-44.

Kwon, Y. J., & A. E. Lawson (2000). Linking brain growth with the development of scientific reasoning ability and conceptual change during adolescence. *Journal of Research in Science Teaching 37*, 44–62.

Lassaline, M. E. (1996). Structural alignment in induction and similarity. *Journal of Experimental Psychology*: *Learning*, *Memory*, *and Cognition*, *22*, 754-770.

Lassaline, M. E., & Murphy, G. L. (1998). Alignment and category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 144-160.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research, 1284,* 145-155.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*, 1077–1083.

Markman, A. B. (1997). Constraints on analogical inference. *Cognitive Science*, *21*, 373-418.

Markman, A. B. (1999). *Knowledge representation.* Hillsdale, NJ: Erlbaum.

Markman, A., & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Science*, *4*, 470–475.

Markman, A. B. and Gentner, D. (1993a). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431- 467.

Markman, A. B., & Gentner, D. (1993b). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*, 517-535.

Markman, A. B., & Gentner, D. (2000). Structure-mapping in the comparison process. *American Journal of Psychology*, *113*, 501-538.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin, 129*, 592-615.

Mason, R. A., & Just, M. A. (2016). Neural representations of physics concepts. *Psychological Science, 27*, 904–913.

Masson, S., Potvin, P., Riopel, M., & Brault Foisy, L.-M. (2014). Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity. *Mind, Brain, and Education, 8*, 44-55.

Masson, S., Potvin, P., Riopel, M., Brault Foisy, L.-M., & Lafortune, S. (2012). Using fMRI to study conceptual change: How and why? *International Journal of Environmental and Science Education, 7*, 19-35.

McElree, B., & Nordlie, J. (1999). Literal and figurative interpretations are computed in equal time. *Psychonomic Bulletin & Review*, *6*, 486-494.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270-291. doi: 10.1037/a0028228

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, E. K., Nieder, A., Freedman, D. J., & Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, *13*, 198-203.

Mitchell, M. & Hofstadter, D. R. (1990). The emergence of understanding in a computer model of concepts and analogy-making. *Physica D, 42*, 322-334.

Morrison, C. T., & Dietrich, E. (1995). Structure-mapping vs. high-level perception: The mistaken fight over the explanation of analogy. *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society*, 678-682.

Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., & Knowlton, B. J., (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *Journal of Cognitive Neuroscience, 16*, 260–271.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289-316.

Nelson, J. K., Lizcano, R. A., Atkins, L., & Dunbar, K. (2007). Conceptuel judgments of expert vs. novice chemistry students: An fRMI study. Paper presented at the 48th Annual meeting of the Psychonomic Society, Long Beach, California.

Nolde, S. F., Johnson, M. K., & D'Esposito, M. (1998). Left prefrontal activation during episodic remembering: An event-related fMRI study. *NeuroReport, 9*, 3509–3514.

Noonan, K. A., Jefferies, E., Visser, M., & Lambon Ralph, M. A. (2013). Going beyond inferior prefrontal involvement in semantic control: Evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *Journal of Cognitive Neuroscience, 25*, 1824–1850.

Norman, D. A., & Rumelhart, D. E. (1975). Memory and knowledge. In D. A. Norman, D. E. Rumelhart, & the LNR Research Group (Eds.), *Explorations in cognition*. San Francisco: Freeman.

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

Ortony, A. (1979). Beyond literal similarity. *Psychological Review*, *86*, 161-180.

Penn, D. C., Holyoak, K J., & Povinelli, D. J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109-178.

Ramnani, N., & Owen, A.M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience, 5*, 184-194.

Ranganath, C., Johnson, M. K., D'Esposito, M. (2000). Left anterior prefrontal activation increases with demands to recall specific perceptual information. *Journal of* Neuroscience, *20*, 108.

Reed, S. (1989). Constraints on the abstraction of solutions. *Journal of Educational Psychology*, *81*, 532-540.

Rehder, B. & Ross, B.H. (2001). Abstract coherent concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1261-1275.

Richardson, J.T. (2007). Measures of short-term memory: A historical review. *Cortex*, *43*, 635–650.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 629-639.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Rozenblit, L. R. & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science, 26,* 521-562.

Schoenbaum, G., Setlow, B. (2001). Integrating orbitofrontal cortex into prefrontal theory: Common processing themes across species and subdivisions. *Learning and Memory*, *8*, 134–47.

Schwartz, D. L. & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*, 475-522. Simon, H. A. (1974). How big is a chunk? *Science*, *183*, 482–88.

Shank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. *The Psychology of Learning and Motivation*, *36*, 45-81.

Simon, H. A. (1974). How big is a chunk? *Science*, *183*, 482–88.

Spellman, B. A., & Holyoak, K. J. (1992). If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles. *Journal of Personality and Social Psychology, 62,* 913-933.

Spellman, B. A., & Holyoak, K. J. (1996). Pragmatics in analogical mapping. *Cognitive Psychology, 31*, 307-346.

Stich, S. P. (1975). *Innate ideas*. Berkley, CA: University of California Press.

Tartter, V. C., Gomes, H., Dubrovsky, B., Molholm, S., & Stewart, R. V. (2002). Novel metaphors appear anomalous at least momentarily: Evidence from N400. *Brain Language*, *80*, 488–509.

Thompson, R. K. R., Oden, D. L. & Boysen, S. T. (1997). Language-naive chimpanzees *(Pan troglodytes) judge* relations be- tween relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal behavior processes, 23,* 31-43.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, USA, 94.* 14792-14797.

Tversky, A. (1977). Features of similarity. *Psychological Reviews 84*, 327–352.

Wagner, A. D., Maril, A., Bjork, R. A. & Schacter, D. L. (2001). Prefrontal contributions to executive control: fMRI evidence for functional distinctions within lateral prefrontal cortex. *NeuroImage*, *14*, 1337–47.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*, 1-39.

Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., & Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": Investigating the role of rostrolateral prefrontal cortex in relational reasoning. *Journal of Cognitive Neuroscience, 20,* 682–693.

Wharton C.M., Grafman J., Flitman, S.S., Hansen, E.K., Brauner, J., & Marks, A., Honda, M. (2000). Toward neuroanatomical models of analogy: a positron emission tomography study of analogical mapping. *Cognitive Psychology, 40,*173-197.

White, I. M, & Wise, S. P. (1999). Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research, 126*, 315–335.

Yamauchi, T., & Markman, A. B. (2000). Inference using categories. Journal of Experimental Psychology: Learning, *Memory and Cognition, 26*, 776-795.

Appendix A: Experiment 1 Scenarios Used in Mapping Condition

Tradeoff

Scenario 1

Jim is looking for a secure investment that will produce a large profit within a few weeks, and which will also produce a steady source of income over several years. Jim can invest in one of two companies. For one of the companies, there is a high probability that investing in it can lead to a large monetary gain over the short term, but the probability is also high that the company will go bankrupt within a few months. For the other company, there is a low probability that it will produce a large profit over the short term, but there is a high probability that it will produce a moderate and steady profit over the course of many years. Jim decides to invest in the second company.

Scenario 2

David is charged with embezzlement. If convicted, he faces a maximum of 5 years in prison. David wants to be acquitted of the charge, but also minimize his risk of serving a prolonged prison sentence. David can either plead guilty, not guilty, or try to work out a plea deal. If the case goes to trial there is a moderate chance that he will be acquitted, but if convicted he will likely be given a minimum 3-year prison sentence. The evidence against David is somewhat strong, but there are some weaknesses in the prosecution's case. David is offered a plea deal in which the burglary charge will be dropped if he agrees to complete 300 hours of community service. David decides to accept the plea deal.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<div style="display: flex; justify-content: space-between;">
<div>

Scenario 1

A. Jim is looking for a secure investment that will produce a large profit within a few weeks, and which will also produce a steady source of income over several years.

B. For one of the companies, there is a high probability that investing in it can lead to a large monetary gain over the short term, but the probability is also high that the company will go bankrupt within a few months. For the other company, there is a low probability that it will produce a large profit over the short term, but there is a high probability that it will produce a moderate and steady profit over the course of many years.

C. Jim can invest in one of two companies.

</div>
<div>

Scenario 2

D. David wants to be acquitted of the charge, but also minimize his risk of serving a prolonged prison sentence.

E. David can either plead guilty, not guilty, or try to work out a plea deal.

F. If the case goes to trial there is a moderate chance that he will be acquitted, but if convicted he will likely be given a minimum 3-year prison sentence.

</div>
</div>

Showdown

Scenario 1

The Cyclops and the Stallions are baseball teams in the Bubblegum Minor Leagues. Both teams have championship aspirations this year. The two teams have played multiple times throughout the season and each game was more contentious than the last. The two teams are currently playing one another in a best of seven playoff series, which stands tied at 3 games apiece. A decisive game 7 will be played tonight, with the winner advancing to the championship and the loser's season coming to an end.

Scenario 2

For many years now, the two opposing political parties have disagreed on a controversial banking regulation. One party wants to keep the regulation in place and the other party wants the regulation removed. After a series of legal battles, the issue is now being reviewed by the Supreme Court. Each party has assembled their best legal team to make a case for their party's side at the trial. After the court's decision, the matter closed and will move on to other matters.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

Scenario 1

A. A decisive game 7 will be played tonight, with the winner advancing to the championship and the loser's season coming to an end.

B. Both teams have championship aspirations this year. The two teams are currently playing one another in a best of seven playoff series, which stands tied at 3 games apiece.

C. The Cyclops and the Stallions are baseball teams in the Bubblegum Minor Leagues.

Scenario 2

D. After the court's decision, the matter closed and will move on to other matters.

E. For many years now, the two opposing political parties have disagreed on a controversial banking regulation.

F. One party wants to keep the regulation in place and the other party wants the regulation removed. After a series of legal battles, the issue is now being reviewed by the Supreme Court.

Risk

Scenario 1

The 54th platoon has been protecting villagers from a rebel insurgency. The villagers want to be sheltered from the fighting and the soldiers want to keep the villagers safe. The platoon has reduced their forces because they are running low on equipment. These factors helped the rebels break through the first two lines of the platoon's infantry. This development has increased the chances that the soldiers and villagers will be harmed.

Scenario 2

Carl is the CEO of a leading tech company. Carl has made numerous poor investments that have cost the company millions, which have caused the company's stock to plummet. The stockholders are hopeful that the company's stocks will stabilize and eventually lead to profits. However, there is a high probability that the company will declare bankruptcy and the stockholders will lose their investments.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<table>
<tr><td align="center">Scenario 1</td><td align="center">Scenario 2</td></tr>
<tr>
<td valign="top">A. The platoon has reduced their forces because they are running low on equipment. These factors helped the rebels break through the first two lines of the platoon's defenses.</td>
<td valign="top">D. However, there is a high probability that the company will declare bankruptcy and the stockholders will lose their investments.</td>
</tr>
<tr>
<td valign="top">B. The villagers want to be sheltered from the fighting and the soldiers want to keep the villagers safe.</td>
<td valign="top">E. Carl is the CEO of a leading tech company. Carl has made numerous poor investments that have cost the company millions, which have caused the company's stock to plummet.</td>
</tr>
<tr>
<td valign="top">C. This development has increased the chances that the soldiers and villagers will be harmed.</td>
<td valign="top">F. The stockholders are hopeful that the company's stocks will stabilize and eventually lead to profits.</td>
</tr>
</table>

Obstruction

Scenario 1

The local bank has been robbed 4 times in the past 3 months and decided to hire Rory the famous bank security guard who has thwarted 15 bank robberies in his prestigious career. Just a few weeks after Rory was hired, a group of bank robbers were staking out the bank to look for weaknesses that they could exploit. However, the would-be robbers were deterred from robbing the bank once they saw Rory.

Scenario 2

A giant dam has recently been built in a small town in California to prevent flooding. There has been heavy rain all day and the height of the river is rising quickly making the locals a bit worried. The current of the river has also increased to an unsettling speed, but fortunately the dam is able to keep the town safe by keeping the water from the river from overflowing into the town.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<table>
<tr><td>Scenario 1</td><td>Scenario 2</td></tr>
<tr><td>A. Just a few weeks after Rory was hired, a group of bank robbers were staking out the bank to look for weaknesses that they could exploit.</td><td>D. A giant dam has recently been built in a small town in California to prevent flooding.</td></tr>
<tr><td>B. However, the would-be robbers were deterred from robbing the bank once they saw Rory.</td><td>E. There has been heavy rain all day and the height of the river is rising quickly making the locals a bit worried.</td></tr>
<tr><td>C. The local bank has been robbed 4 times in the past 3 months and decided to hire Rory the famous bank security guard who has thwarted 15 bank robberies in his prestigious career.</td><td>F. ...fortunately the dam is able to keep the town safe by keeping the river from overflowing into the town.</td></tr>
</table>

Investigation

Scenario 1

A health committee was recently appointed by group of neighborhood businesses. The committee has been asked to look into whether the new bar is being sanitary. Throughout the following week, the health committee sends undercover agents to keep tabs on the bar.

Scenario 2

Kelly is a physicist at the local university. Kelly is interested in studying how high levels of heat affects subatomic particles. She conducts a series of experiments in the laboratory where she manipulates the temperature and then collects the subatomic particle readings.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<table>
<tr><td align="center">Scenario 1</td><td align="center">Scenario 2</td></tr>
<tr><td>A. A health committee was recently appointed by group of neighborhood businesses.</td><td>D. Kelly is interested in studying how high levels of heat affects subatomic particles.</td></tr>
<tr><td>B. The committee has been asked to look into whether the new bar is being sanitary.</td><td>E. Kelly is a physicist at the local university.</td></tr>
<tr><td>C. Throughout the following week, the health committee sends undercover agents to keep tabs on the bar.</td><td>F. She conducts a series of experiments in the laboratory where she manipulates the temperature and then collects the subatomic particle readings.</td></tr>
</table>

Deception

Scenario 1

The manager at a local restaurant is under investigation for serving expired meat. Health inspectors came by the restaurant to examine the meat. However, the day before they arrived, the restaurant's manager temporarily stored all of the expired meat in a different location and only showed the inspectors the new meat. When asked if there was any other meat in the restaurant, the restaurant's manager told the inspectors that there was not. Satisfied with their findings, the inspectors dismissed the claims against the restaurant.

Scenario 2

The general's army was moments from losing the battle so the general deiced to surrender. The opposing leader came out to discuss the general's terms of surrender. However, this was a diversion strategy by the general to buy time for his troops. The general had called for reinforcements the week before, which were due to arrive within the hour. Caught off guard by the reinforcements, the opposing army was swiftly defeated and the major and his troops survived the battle.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

Scenario 1

A. However, the day before they arrived, the restaurant's manager temporarily stored all of the expired meat in a different location and only showed the inspectors the new meat. When asked if there was any other meat in the restaurant, the restaurant's manager told the inspectors that there was not.

B. The manager at a local restaurant is under investigation for serving expired meat.

C. Health inspectors came by the restaurant to examine the meat.

Scenario 2

D. The general's army was moments from losing the battle so the general deiced to surrender.

E. The opposing leader came out to discuss the general's terms of surrender.

F. However, this was a diversion strategy by the general to buy time for his troops. The general had called for reinforcements the week before, which were due to arrive within the hour.

Cooperation

Scenario 1

Over the past decade, pollution has become a major concern. To help combat this issue, Brazil, Sweden, and Finland have agreed to work together on a special project. The aim of the project is to cut the world's pollution by 30 percent within the next 5 years. Brazil has volunteered its top scientists, Sweden has volunteered its top engineers, and Finland has volunteered its top computer scientists to work on the project. All three countries will collect and share data with one another and meet to discuss a joint solution.

Scenario 2

Sam and Bill work for the coastguard and were separated from their group in a training exercise over the pacific. They swim to a deserted island, where they must survive until a rescue crew can find them. A storm is coming in and it will be at least 10 days before a rescue crew arrives. During the day, Sam works on a way to make the ocean water drinkable and Bill works on finding food. Sam and Bill share their goods to help ensure they survive.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<div align="center">Scenario 1</div>

A. The aim of the project is to cut the world's pollution by 30 percent within the next 5 years.

B. Brazil has volunteered its top scientists, Sweden has volunteered its top engineers, and Finland has volunteered its top computer scientists to work on the project. All three countries will collect and share data with one another and meet to discuss a joint solution.

C. ...Brazil, Sweden, and Finland have agreed to work together on a special project.

<div align="center">Scenario 2</div>

D. They swim to a deserted island, where they must survive until a rescue crew can find them.

E. During the day, Sam works on a way to make the ocean water drinkable and Bill works on finding food. Sam and Bill share their goods to help ensure they survive.

F. Sam and Bill work for the coastguard and were separated from their group in a training exercise over the pacific.

Compensation

Scenario 1

Johnny works as a cashier at the local supermarket. Last week, one of the market's employees improperly stacked some boxes in the backroom. The boxes ended up falling on Johnny, giving him a concussion. As a result, the manager of the supermarket decided to award Johnny 30 thousand dollars to cover Johnny's medical bills and the pain and suffering.

Scenario 2

A fight broke out in front of Karen and Tom during halftime at a basketball game. In the midst of the scuffle, a soda was spilled on Tom's coat. Karen raised these issues in a letter she wrote to the home team's owner. The team owner responded by giving Karen and Tom courtside seats for the remainder of the season.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<table>
<tr><td align="center">Scenario 1</td><td align="center">Scenario 2</td></tr>
<tr><td>A. Johnny works as a cashier at the local supermarket.</td><td>D. In the midst of the scuffle, a soda was spilled on Tom's coat.</td></tr>
<tr><td>C. ...the manager of the supermarket decided to award Johnny 30 thousand dollars to cover Johnny's medical bills and the pain and suffering.</td><td>E. A fight broke out in front of Karen and Tom during halftime at a basketball game.</td></tr>
<tr><td>D. Last week, one of the market's employees improperly stacked some boxes in the backroom. The boxes ended up falling on Johnny, giving him a concussion.</td><td>F. The team owner responded by giving Karen and Tom courtside seats for the remainder of the season.</td></tr>
</table>

Aid

Scenario 1

Robert broke his leg in a boating accident 1 week ago and it is still difficult for him to get around the house. Robert has mostly been staying at home with his trusty dog Max. Robert is watching television on his couch and wants to change the channel, but left the remote control on his coffee table. It is still very painful for Robert to stand once he has been sitting so he does not want to walk over to the coffee table. Max runs to the coffee table and brings the remote control to Robert. Robert is now able change the channel without having to get up off the couch.

Scenario 2

Tom is enrolled in a calculus course and has done poorly on the previous 2 exams and needs to score an 82 on the final to pass the course. However, Tom does not understand some of the material that was covered in the past 7 weeks. Marcy, the top student in the course, recognizes that Tom is struggling and offers to tutor him the week before the final. For one week, Marcy tutors Tom for 3 hours a day. Marcy is able to explain the material in a way that makes sense to Tom. Tom ends up scoring a 97 on the final and passes the course.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<table>
<tr><td align="center">Scenario 1</td><td align="center">Scenario 2</td></tr>
<tr><td valign="top">A. Max runs to the coffee table and brings the remote control to Robert.</td><td valign="top">D. Tom is enrolled in a calculus course and has done poorly on the previous 2 exams and needs to score an 82 on the final to pass the course.</td></tr>
<tr><td valign="top">B. Robert is now able change the channel without having to get up off the couch.</td><td valign="top">E. Marcy, the top student in the course, recognizes that Tom is struggling and offers to tutor him the week before the final.</td></tr>
<tr><td valign="top">C. Robert is watching television on his couch and wants to change the channel, but left the remote control on his coffee table. It is still very painful for Robert to stand once he has been sitting so he does not want to walk over to the coffee table.</td><td valign="top">F. Marcy is able to explain the material in a way that makes sense to Tom. Tom ends up scoring a 97 on the final and passes the course.</td></tr>
</table>

Reciprocity

Scenario 1	Scenario 2
Becky and Dave went on a trip where they visited a village in the amazon. As a welcoming gift, the village tribe gave Becky and Dave an assortment of rare fruits that are native to the island. To show their appreciation, Becky and Dave handcrafted bracelets for each of the villagers.	Due to budget cuts, the old town orphanage is set to close at the end of the month. Prompted by the news, a local millionaire donated 5.5 million dollars to the orphanage so that it could remain open. As a gesture of goodwill, the workers at the orphanage sent the millionaire 3-dozen boxes of homemade cookies.

Below are three passages from each scenario. For each passage in Scenario 1 (A-C), indicate which passage in Scenario 2 (D-F) it is most similar or analogous to (based on the type of information they convey). The passage labels from Scenario 1 are presented below and each has a textbox next to it. For each of these passages, type in the corresponding letter option from Scenario 2 into its textbox. For example, if Passage A is most similar to Passage F, you would type "F" into the textbox next to the Passage A label.

<div style="display:flex">
<div>

Scenario 1

A. To show their appreciation, Becky and Dave handcrafted bracelets for each of the villagers.

B. Becky and Dave went on a trip where they visited a village in the amazon.

C. As a welcoming gift, the village tribe gave Becky and Dave an assortment of rare fruits that are native to the island.

</div>
<div>

Scenario 2

D. Prompted by the news, a local millionaire donated 5.5 million dollars to the orphanage so that it could remain open.

E. Due to budget cuts, the old town orphanage is set to close at the end of the month.

F. As a gesture of goodwill, the workers at the orphanage sent the millionaire 3-dozen boxes of homemade cookies.

</div>
</div>

Appendix B: Experiments 2-4 Stimuli

Morkels	Non-Morkels
1. Operates on land Works to gather harmful solids Has a shovel	1. Operates on land Works to clean spilled oil Has an electrostatic filter
2. Operates on the surface of the water Works to clean spilled oil Has a spongy material	2. Operates on the surface of the water Works to collect dangerous gaseous ions Has a shovel
3. Operates in the stratosphere Works to collect dangerous gaseous ions Has an electrostatic filter	3. Operates in the stratosphere Works to gather harmful solids Has a spongy material
4. Operates in highway tunnels Works to remove carbon dioxide Has a large intake tank	4. Operates in highway tunnels Works to remove lost fishing nets Has a sifter
5. Operates in swamps Works to remove malaria-ridden mosquitos Has a finely woven net	5. Operates in swamps Works to remove broken glass Has a metal pole with sharpened end
6. Operates in warzones Works to gather shards of metal Has a large magnet	6. Operates in warzones Works to gather discarded paper Has a finely woven net
7. Operates in parks Works to gather discarded paper Has a metal pole with a sharpened end	7. Operates in parks Works to gather shards of metal Has a hook
8. Operates on the seafloor Works to remove lost fishing nets Has a hook	8. Operates on the seafloor Works to remove malaria-ridden mosquitoes Has a large intake fan
9. Operates on the beach Works to remove broken glass Has a sifter	9. Operates on the beach Works to remove carbon dioxide Has a large magnet
10. Operates on wood floors Works to remove stains Has absorbent cloth	10. Operates on wood floors Works to collect ocean sediments Has an intake port

Morkels cont'd	Non-Morkels cont'd
11. Operates on solid surfaces Works to remove debris Has a dense set of bristles	11. Operates on solid surfaces Works to remove large rocks Has absorbent cloth
12. Operates on glass Works to remove liquids Has a rubber edge	12. Operates on glass Works to collect small particles Has a dense set of bristles
13. Operates on thick woven fabric Works to collect small particles Has an intake port	13. Operates on thick woven fabric Works to remove brush Has a large shovel
14. Operates in harbors Works to collect ocean sediments Has a large shovel	14. Operates in harbors Works to remove leaves Has metal teeth and a sieve
15. Operates in the jungle Works to remove brush Has sharp blades	15. Operates in the jungle Works to remove debris Has a rubber edge
16. Operates in farmland Works to remove rocks Has metal teeth and a sieve	16. Operates in farmland Works to smooth rough spots Has a rough metal surface
17. Operates on fine wood Works to smooth rough spots Has a rough metal surface	17. Operates on fine wood Works to remove liquids Has a nozzle and a motor
18. Operates in gardens Works to remove leaves Has a nozzle and a motor	18. Operates in gardens Works to remove stains Has sharp blades

Note: Morkels were coherent items and non-Morkels were incoherent items. The first three items in each category were taken from Rehder and Ross (2001), the next six items were taken from Higgins (2012), and the remaining items were created by the present authors.

Appendix C: Experiment 6 Stimuli

Zorpes	Inference Lure 1	Inference Lure 2
1. Dog Car Asteroid	1. Train Hamster Chocolate Bar	1. Subway Train Mouse Peanut
2. Pocket Knife Book Television	2. Door Mosquito Fingernail	2. Apartment Caterpillar Eyelash
3. Desk Lamp Bicycle Bear	3. Hippopotamus Stapler Paperclip	3. Great White Shark Sunglasses Light Bulb
4. Apple Cat Boat	4. Dolphin Worm Dollar Bill	4. Panda Moth Goldfish
5. Remote Control Laptop Desk	5. Sofa Safety Pin Wasp	5. Gorilla Credit Card Grasshopper
6. Pencil Shoe Tiger	6. Bookshelf Electron Ladybug	6. Skateboard Beetle AA Batteries
7. Wallet Coffee Table Lion	7. Buffalo Key Glove	7. Mattress Cricket Backpack
8. Banana A Volleyball Penguin	8. Wrecking Ball Bacteria A Baseball	8. Missile Tooth Grain of Sand
9. Ring Mango Dragon	9. Watermelon Atoms Coin	9. Refrigerator Pea Thumb
10. Nail Polish Container Glasses Hat	10. Window DNA Molecules Flash Drive	10. Lawn Mower Flea Chap Stick

Zorpes cont'd	Inference Lure 1 cont'd	Inference Lure 2 cont'd
11. Salt Shaker Hammer Whale	11. Stove Sewing Pin Spoon	11. Shovel Grain of Salt Snail
12. Teddy Bear Kangaroo Building	12. Tank Lipstick Pie	12. Submarine Nail Pizza Slice
13. Chair Truck Saturn	13. Tugboat Plate Manikin	13. Parking Lot Baseball Mitt Typewriter
14. Kitchen Trey Elephant Mountain	14. Moon Keychain Deer	14. Sequoya Tree Lighter Carrot
15. Eraser Hamburger Tree	15. Car Battery Microbe Ant	15. Swimming Pool Sesame Seed Slug
16. Hippopotamus Cruise Ship Jupiter	16. Sun Garden Snake Paddle	16. Mount Fuji Possum Purse
17. Crocodile Tyrannosaurus Rex Statue of Liberty	17. Venus Goose Cabinet Drawer	17. Skyscraper Skunk Piano
18. Rhinoceros Jumbo Jet The Great Pyramids	18. China Windshield Belt	18. Empire State Building Fox Pillow

Olatin	Inference Lure 1	Inference Lure 2
1. Comet Volkswagen Beetle Wolf	1. Bowling Ball Solar System Helicopter	1. Bee Neptune Aircraft Carrier
2. Movie Theater Raft Bucket	2. A Crown Grand Canyon Eifel Tower	2. Rooster National Park Bulldozer
3. Garage Motorcycle Butterfly	3. Hotdog Fort Knox Spaceship	3. Trout Rainforest Cement Mixer
4. Lighthouse Octopus Almonds	4. Rabbit Black Hole Battleship	4. Squirrel Pluto Waterfall
5. Jaguar Ironing Board Nail Clipper	5. Grocery Bag Mall Palm Tree	5. Toothbrush Lincoln Memorial Ice Rink
6. Porch Deck Bed Steering Wheel	6. Dustpan Museum Arena	6. Violin Gymnasium Water Park
7. Panther Vacuum Cleaner Hair Comb	7. Headphones Amusement Park Sears Tower	7. Microscope Tow Truck Palace
8. Mars Waterslide Bench	8. Ostrich Galaxy Stadium	8. Strawberry Universe Mount Everest
9. Blimp Puma Magnifying Glass	9. Cereal Bowl Mount Rushmore Hut	9. Chestnut Niagara Falls Hotel
10. Alligator Kitchen Sink Fish Bowl	10. Chopsticks Triceratops Military Base	10. Chipmunk Manmade Dam Pine Tree

Olatin cont'd	Inference Lure 1 cont'd	Inference Lure 2 cont'd
11. Saber-Toothed Tiger Sword Doorknob	11. Tennis Ball Pacific Ocean Supermarket	11. Cherry Coliseum Casino
12. Ferris Wheel Washing Machine Coconut	12. Slice of Bread Swamp Football Field	12. Camera Constellation Oil Rig
13. Oven DVD Player Soda Can	13. Bottle Cap Bar Filing Cabinet	13. Laser Pointer Cabin Catapult
14. Horse Eagle Frisbee	14. Parrot Farm School Bus	14. Contact Lens Concert Hall Capitol Building
15. Pool Table Binoculars Bar of Soap	15. Ice-Cream Cone Ambulance Cougar	15. Pebble Bowling Alley Satellite
16. Recliner Baseball Bat Bowling Pin	16. Crayon Iceberg Jeep	16. Toothpick Hospital Crane
17. Bull Microwave Pineapple	17. Golf Ball Garbage Truck Polar Bear	17. Test Tube Mansion Forklift
18. Cannon Duck Pepper Spray	18. Egg Airport Sailboat	18. Scissors Fire Truck Space Station

Note: Each lure pertains to the stimulus' component on the corresponding row.