Restructuring partitioned knowledge: The role of recoordination in category learning☆

David K. Sewell[*,a,b], Stephan Lewandowsky[b]

[a] *The University of Melbourne*
[b] *The University of Western Australia*

Abstract

Knowledge restructuring refers to changes in the strategy with which people solve a given problem. Two types of knowledge restructuring are supported by existing category learning models. The first is a relearning process, which involves incremental updating of knowledge as learning progresses. The second is a recoordination process, which involves novel changes in the way existing knowledge is applied to the task. Whereas relearning is supported by both single- and multiple-module models of category learning, only multiple-module models support recoordination. To date, only relearning has been directly supported empirically. We report two category learning experiments that provide direct evidence of recoordination. People can fluidly alternate between different categorization strategies, and moreover, can reinstate an old strategy even after prolonged use of an alternative. The knowledge restructuring data are not well fit by a single-module model (ALCOVE). By

☆This paper is based on a doctoral dissertation submitted to The University of Western Australia by the first author. Preparation of this paper was facilitated by an Australian Postgraduate Award Scholarship to the first author and a Discovery Grant from the Australian Research Council and an Australian Professorial Fellowship to the second author. We gratefully acknowledge the support of Philip Smith during the latter stages of this project. We thank Stewart Craig, Charles Hanich, and Leo Roberts for their assistance with data collection. We thank Donna Bayliss, James Dixon, Arthur Markman, and two anonymous reviewers for valuable comments on a previous version of this paper.

[*]Address correspondence to David K. Sewell, Psychological Sciences, The University of Melbourne, VIC 3010, Australia

Email addresses: `dsewell@unimelb.edu.au` (David K. Sewell), `lewan@psy.uwa.edu.au` (Stephan Lewandowsky)

URL: `http://www.psych.unimelb.edu.au/people/staff/SewellD.html` (David K. Sewell), `http://www.cogsciwa.com` (Stephan Lewandowsky)

contrast, a multiple-module model (ATRIUM) quantitatively accounts for recoordination. Low-level changes in the distribution of dimensional attention are shown to subsequently affect how ATRIUM coordinates its modular knowledge. We argue that learning about complex tasks occurs at the level of the partial knowledge elements used to generate a response strategy.

Key words: Knowledge Restructuring, Knowledge Partitioning, Strategy Change, Category Learning, Categorization, Computational Modeling, Mixture of Experts

---

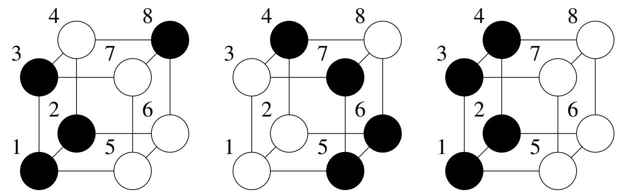
1. Introduction

A fundamental question in cognitive science concerns the interplay between the structure of existing knowledge and the acquisition of new knowledge. Typically, this question has been explored by examining a) how existing knowledge constrains the acquisition of new knowledge, or b) how the acquisition of new knowledge alters the structure of existing knowledge. Both approaches have revealed that much of learning and development involves qualitative changes in the structure, or application, of domain-specific knowledge. For example, young children often approach arithmetic problems (e.g., $3 + 5 = ?$) by counting with their fingers (Shrager & Siegler, 1998; Siegler, 1987; Siegler & Jenkins, 1989) whereas adults typically rely on retrieval of the solution from memory (Delaney, Reder, Staszewski, & Ritter, 1998; Rickard, 1997; Staszewski, 1988). We refer to the empirically identifiable emergence of novel strategies in the absence of any change in task environment as *knowledge restructuring* (Kalish, Lewandowsky, & Davies, 2005). That a stable task environment is a defining property of knowledge restructuring permits contrast with situations in which novel strategies emerge in response to changes in task structure; we refer to the latter as *strategy change*. The emergence of novel strategies distinguishes knowledge restructuring and strategy change from other forms of incremental learning, which involve the refinement of an existing strategy (e.g., what Kalish et al., 2005 refer to as performance improvement).

In this paper, we adopt a model-based analysis of the linkage between the structure of category knowledge on the one hand, and processes of knowledge restructuring on the other. Our principal argument is that different functional architectures support different types of knowledge restructuring. Our theoretical aim is to determine the extent to which computational principles underpinning existing category learning models can account for different forms of knowledge restructuring.

Our focus on categorization is motivated by several reasons. First, because knowledge restructuring is ostensibly domain general (Carey, 1985, 1991, 2009; Chase & Simon, 1973a,b; Chi, Feltovich, & Glaser, 1981; Dixon & Bangert, 2002; Dixon & Dohn, 2003; Gopnik & Meltzoff, 1997; Johnson & Carey, 1998; Keil, 1989; Shafto & Coley, 2003; Siegler, 1981, 1987, 1996; Siegler & Jenkins, 1989; Vosniadou, 1994; Vosniadou & Brewer, 1992, 1994), an exhaustive analysis is beyond the scope of a single paper; we restrict our analysis accordingly. Second, categorization affords a high degree of experimental control over what is learned because the category structure and stimuli are determined in advance. Third, generalization to novel stimuli presented after training readily reveals people's response strategies. Finally, a focus on categorization foregrounds the issue of representational structure. Category representations have been described in terms of rules (Ashby & Gott, 1988; Ashby & Townsend, 1986; Bruner, Goodnow, & Austin, 1956), prototypes (Posner & Keele, 1968; Rosch, 1975; Smith & Minda, 1998), collections of exemplars (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986; Nosofsky & Johansen, 2000), clusters (Anderson, 1991; Love, Medin, & Gureckis, 2004), or combinations of the above, the most common being a rule-plus-exemplar approach (Erickson & Kruschke, 1998; Nosofsky, Clark, & Shin, 1989; Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994). A number of these views have been developed into computational models (see Kruschke, 2008, for a recent review), which permits quantitative assessment of knowledge restructuring and strategy use.

Given the prevalence of category learning models, it is perhaps surprising that only a handful of studies have used such models to examine knowledge restructuring and strategy change (e.g., Johansen & Palmeri, 2002; Kruschke, 1996; Little, Lewandowsky, & Heit, 2006; Macho, 1997). A model-based approach is particularly illuminating because different model architectures permit different processes of knowledge restructuring that may (or may not) be used by human learners. For simplicity, we focus on the distinction between single-module and multiple-module architectures, limiting our analysis to the connectionist framework within which Kruschke and colleagues have developed both single- and multiple-module models. Single-module models, such as ALCOVE (Kruschke, 1992), assume that a unitary underlying knowledge structure classifies all stimuli. By contrast, multiple-module models assume a heterogeneous knowledge structure, which permits application of only a sub-



(a) Type V (b) Complete Reversal (c) Exception Reversal

Figure 1: Logical structure of the problems used to illustrate two forms of knowledge restructuring. Category assignments of the eight stimuli are represented by filled (Category A) and unfilled (Category B) circles. The Type 5 problem is shown in the leftmost panel. The middle panel shows category assignments in the Complete Reversal simulation. The right panel shows category assignments in the Exception Reversal simulation.

set of category knowledge on a given trial. For example, models that store rules as well as exemplars in memory (Erickson & Kruschke, 1998; Nosofsky et al., 1994; Nosofsky & Palmeri, 1998) may employ a rule on some trials, but rely on exemplar memory on others.

Functional architecture carries important implications for knowledge restructuring. Single-module architectures, because they possess a unitary knowledge structure, principally rely on an incremental form of knowledge restructuring, which we refer to as *relearning*. Relearning is readily described in terms of incremental adjustments to a single associative weight matrix. A novel response strategy emerges once sufficient incremental change has been accumulated. Multiple-module models are more flexible and can support at least two varieties of restructuring: the relearning process available to single-module models, and additionally, a *recoordination* process that can describe abrupt changes in response strategy. Recoordination involves changes in the way existing subsets of knowledge are applied to the task (e.g., a rule might be applied to a stimulus previously categorized using an exemplar strategy). Interestingly, although there is broad empirical support for a relearning process, very little direct evidence has been adduced for recoordination. The major empirical contribution of this paper is a direct verification of recoordination-based knowledge restructuring.

Before reviewing the relevant empirical literature, we summarize novel simulation results using ALCOVE and ATRIUM to illustrate the differences between relearning and recoordination. To facilitate exposition, we provide a brief overview of the two models. The models are presented formally in the Computational Modeling section of this article.

2. Two Forms of Knowledge Restructuring

To illustrate the relationship between relearning, recoordination, and model architecture, we report two novel simulations applying ALCOVE and

ATRIUM to the Type 5 category structure introduced by Shepard, Hovland, and Jenkins (1961). The Type 5 structure comprises three binary-valued stimulus dimensions. The stimuli, defined by all possible (i.e., 2^3) combinations of stimulus values, are divided equally into two categories (see panel A of Figure 1). We chose this particular structure for two reasons: First, ALCOVE has been shown to learn all of the Shepard problems, including the Type 5 structure (Kruschke, 1992). Second, the Type 5 structure is amenable to a modular rule-plus-exception strategy to which ATRIUM is particularly suited. The category structure can be characterized by a rule that bisects the x dimension with one exception stimulus per category (Items 4 and 8).

ALCOVE (Kruschke, 1992) operates by storing training exemplars in memory, and learns to associate those exemplars with category responses. Learning is modeled by adjustment of exemplar-to-category association weights to minimize error[1]. At any point in time, ALCOVE's "knowledge" is summarized by the single matrix of association weights between stored exemplars and category responses. ALCOVE learned the Type 5 problem (like all other Shepard problems) by associating each stimulus with its assigned category label.

ATRIUM (Erickson & Kruschke, 1998, 2002a) possesses a modular architecture that minimally combines an exemplar representation (an implementation of ALCOVE) with a rule representation that bisects the category space into two response regions along a single psychological dimension. Learning occurs by adjustment of rule-to-category and exemplar-to-category association weights in the rule and exemplar modules, respectively. Within ATRIUM, each module independently attempts to learn the entire category structure; hence, the modules interact competitively. A gating mechanism mediates competition between modules by learning to associate each training exemplar with the module that most accurately categorizes it. The gating mechanism allows ATRIUM to develop a heterogeneous, stimulus-dependent knowledge structure that is more nuanced than ALCOVE's. Rule-based knowledge can be exclusively applied to one subset of stimuli, whereas exemplar-based knowledge can be applied to another subset. ATRIUM learned a rule-plus-exception solution to the Type 5 problem by using a rule

[1] ALCOVE and ATRIUM share an additional attentional learning mechanism that allows them to selectively attend to categorically diagnostic stimulus dimensions. Because this mechanism is not relevant to the demonstration below, we do not discuss it here.

along the x dimension to assign Items 1, 2, and 3 to Category A, and Items 5, 6, and 7 to Category B. The exceptions (Items 4 and 8) were categorized by the exemplar module. Full details about the simulations, including training routines, model parameters, and learned weight matrices are presented in Appendix A.

After the models had learned the Type 5 structure, we implemented two different types of category reversal shifts (cf. Kruschke, 1996), whereby reinforcement contingencies were reversed for certain subsets of stimuli, and the models were retrained. In the *Complete Reversal* simulation, all category labels were reversed (see panel B in Figure 1). In the *Exception Reversal* simulation, only exception item reinforcement was reversed (panel C in Figure 1). The Exception Reversal is equivalent to reducing the category structure to a Type 1 problem, which can be solved by applying a single (exception-free) rule.

Both models accommodated the reversal shifts (see Appendix A for details). For the Complete Reversal simulation, both models adapted to the reversal shift via a relearning process. In ALCOVE, the strengths of the exemplar-to-category association weights were reversed relative to their initially learned values, reflecting the change in reinforcement. An analogous relearning process occurred within each of ATRIUM's modules. In the rule module, rule-to-category weights were reversed relative to their initially learned values; similar changes for exception item weights occurred in the exemplar module. Crucially, there were no changes at the level of the gating mechanism.

ALCOVE responded to the Exception Reversal in much the same way it responded to the Complete Reversal. Only association weights for the exception items were relearned, as it tracked the changes in reinforcement for those items. ATRIUM, by contrast, did not adapt to the Exception Reversal via a relearning process. Rather than relearning exception item weights in the exemplar module, ATRIUM responded to the reversal shift at the level of the gating mechanism. After the reversal shift, Items 4 and 8, which were previously categorized exclusively by the exemplar module were predominantly categorized by the rule module. Thus, ATRIUM's response to the Exception Reversal shift can be characterized in terms of a shift from a rule-plus-exception strategy to a rule-based strategy. Because the change in behavior arose by expanding the application of existing rule-based knowledge to include Items 4 and 8, we argue that ATRIUM's behavior is best characterized in terms of a recoordination process.

Our simulation results provide a theoretical existence-proof for the process of recoordination. The fact that ALCOVE's behavior could not change in this way tentatively identifies two necessary conditions for recoordination to occur. First, a heterogeneous knowledge structure must underpin task performance. Second, existing knowledge must be commensurate with multiple response strategies. We review relevant studies before further developing the idea of recoordination.

3. Knowledge Restructuring in Category Learning

Knowledge restructuring studies examine changes in performance in the absence of any changes in task structure. Studies differ as to whether restructuring was induced by extended training or explicit hints. Training-induced studies have identified spontaneous changes in response strategy over the course of extended training (e.g., Bourne, Healy, Kole, & Graham, 2006; Bourne, Healy, Parker, & Rickard, 1999; Johansen & Palmeri, 2002). In hint-induced studies, restructuring is experimentally encouraged by providing an explicit hint describing an alternative strategy after some initial training (e.g., Kalish et al., 2005; Lewandowsky, Kalish, & Griffiths, 2000; Little, Lewandowsky, & Heit, 2006).

A typical finding from training-induced studies (Bourne et al., 1999, 2006; Johansen & Palmeri, 2002) is that people restructure from an initial rule-based strategy to a memory-based exemplar strategy later on. In the studies conducted by Bourne et al., all stimuli could be categorized by rules (e.g., Respond "A" if a string of letters can be rearranged to form a consecutive sequence). After classifying each item, participants indicated whether they were guessing, recalling the correct response from a previous trial, or applying a rule. People reported relying on rules in the early stages of learning, but later reported using a retrieval-based strategy.

Johansen and Palmeri (2002) explored knowledge restructuring using the 5-4 category structure introduced by Medin and Schaffer (1978). The 5-4 structure comprises 9 training stimuli, five items in one category, four items in the other, instantiated along four binary dimensions. Within each category, stimuli have a family resemblance, meaning that the majority of their features are typical of other members of their category. Critically, each stimulus has at least one feature that is typical of the contrast category. These exception features are distributed across all nine stimuli such that no unidimensional rule is sufficient to classify the entire training set.[2] Johansen and Palmeri's modeling results were consistent with Bourne et al.'s (1999; 2006) data. Categorization of early transfer stimuli was best fit by a model instantiating simple (but imperfect) uni-dimensional rules, whereas later performance was best fit by an exemplar-based model.

Studies that encouraged knowledge restructuring by providing an explicit hint about an alternative strategy have typically used larger stimulus sets and more complex category structures (e.g., Kalish et al., 2005; Lewandowsky et al., 2000; Little et al., 2006). Those studies have revealed that although a hint is necessary for knowledge restructuring, it is individually insufficient—the current strategy must be associated with some amount of error[3].

In the study by Kalish et al. (2005), participants were trained on a large set of stimuli (144 items) that varied along three dimensions: A dichotomous dimension and two quasi-continuous dimensions. Categorization could be achieved by two strategies: An expedient but imperfect strategy, relying only on the dichotomous dimension, or a more complex strategy utilizing both quasi-continuous dimensions, which could yield perfect performance. The cue validity of the dichotomous dimension was manipulated (ranging from 60% to 100%), thereby altering the level of performance error associated with the expedient strategy. When a hint revealing the complex strategy was provided halfway through training, restructuring only occurred when the expedient strategy was associated with high levels of error. When the expedient strategy generated little error, people resisted restructuring (see also Lewandowsky et al., 2000). Kalish et al. (2005) further showed that the occurrence of restructuring was contingent on the hint. When a hint was withheld, even after completing additional training, and regardless of the level of error associated with the expedient strategy, people did not spontaneously restructure to the complex strategy.

[2]For example, if one of the dimensions described the shape of a stimulus (e.g., a square or a triangle), all of the Category A stimuli, say, bar one, would be squares, whereas all of the Category B stimuli, bar one, would be triangles.

[3]Knowledge restructuring can also occur in the absence of error (Dixon & Bangert, 2002; Dixon & Dohn, 2003). Dixon and colleagues use the term representational redescription to describe restructuring driven by accurate task performance. Redescription occurs when repeated success with a particular strategy disembeds relational information contained in the current strategy, resulting in a more robust abstract representation of the task. We do not consider this form of knowledge restructuring here.

The studies by Kalish, Lewandowsky, and colleagues established the necessary and sufficient conditions for restructuring in response to a hint. However, because their studies always involved a shift from an error-prone to a (potentially) error-free strategy, it remains to be seen whether restructuring between distinct, but equally-effective strategies is governed by the same principles. It also remains unclear whether the process responsible for the observed restructuring is one of relearning or recoordination.

Perhaps the strongest support for a recoordination process comes from a study by Little, et al. (2006), who examined knowledge restructuring of ad hoc categories (categories constructed in response to novel goals, e.g., "Things to take from a house on fire", see Barsalou, 1983, 1985). Little et al. trained people to assign stimuli into 4 different categories. Stimuli were surreptitiously grouped into categories according to different themes (i.e., ad hoc categories) with one anomalous exception per category. Anomalous items fit thematically with one of the ad hoc categories used in the study, but were assigned to a contrast category via training. For example, Jewelry, Documents, and Heirlooms are all examples of "Things to take from a house on fire". Apples, Leaves, and Snow are examples of "Things that can fall on your head". Little et al. trained people to assign Apples, Leaves, and Heirlooms to one category, and Jewelry, Documents, and Snow to another. Note that the first group of items are examples of "Things that can fall on your head", with the exception of the anomalous item *Heirlooms*. Likewise, the second group are all "Things to take from a house on fire" with the exception of *Snow*. Thus, people's pre-experimental knowledge about ad hoc categories was placed in competition with experimentally acquired knowledge for anomalous items. Upon revelation of the ad hoc category labels (i.e., a hint), people restructured their knowledge, changing to a categorization strategy that exploited their existing pre-experimental knowledge about the ad hoc categories. Using the above example, people began to assign *Snow* to the same category as other "Things that can fall on your head" — an assignment that conflicted with training. Little et al. showed via computational modeling that the relationship between pre-experimental knowledge and experimentally acquired knowledge was functionally modular. Under instructions to categorize stimuli according to their trained categories, people tended to rely on pre-experimental knowledge to categorize stimuli after the ad hoc category labels were revealed. The selective application of existing pre-experimental knowledge provides an empirical example of a recoordination process.

There are several differences between the study by Little et al. (2006) and those of Kalish et al. (2005) and Lewandowsky et al. (2000) that question the generality of the recoordination process. First, pre-experimental knowledge was not applicable in the Kalish et al. study. Second, unlike the Kalish et al. and Lewandowsky et al. studies, which involved shifts from an error-prone to a (potentially) error-free strategy, the prior knowledge strategy in the Little et al. study was clearly simpler than the experimentally acquired one. The former involved only thematic relations, whereas the latter involved learning arbitrary item-to-category mappings. Thus, it is unclear whether recoordination is restricted to strategy shifts involving a switch to a simpler strategy that is firmly grounded in pre-experimental knowledge. It is therefore important to control for pre-experimental knowledge, whilst ensuring that restructuring does not only progress from a simple strategy to a complex one. We next present a novel paradigm that can simultaneously satisfy these constraints.

4. Knowledge Partitioning

The earlier ATRIUM simulations established that heterogeneity of category knowledge appears to be a precondition for recoordination to emerge. We thus chose a paradigm that has shown evidence of hetereogeneous learning; namely, the knowledge partitioning framework (Lewandowsky, Roberts, & Yang 2006; Little & Lewandowsky, 2009; Yang & Lewandowsky, 2003, 2004; see also Erickson, 2008). Knowledge partitioning occurs when a normatively irrelevant "context" cue (e.g., the color in which a stimulus is presented) reliably signals the presence of a local regularity in the task environment (e.g., the presence of a partial category boundary). When knowledge partitioning occurs, people simplify a complex task by decomposing it into separate local solutions. This paradigm is central to the present studies; to illustrate, we discuss the category space introduced by Yang and Lewandowsky (2003), shown in Figure 2.

During training, the dichotomous context cue was used to differentiate stimuli clustered around the upper and lower partial boundaries (i.e., stimuli within each cluster were presented in a certain color). Context thus predicted which partial boundary could be applied to classify each stimulus (refer to the training stimuli enclosed by the circles in Figure 2) without being diagnostically relevant on its own (i.e., an equal number of Category A and B items were presented in each context).

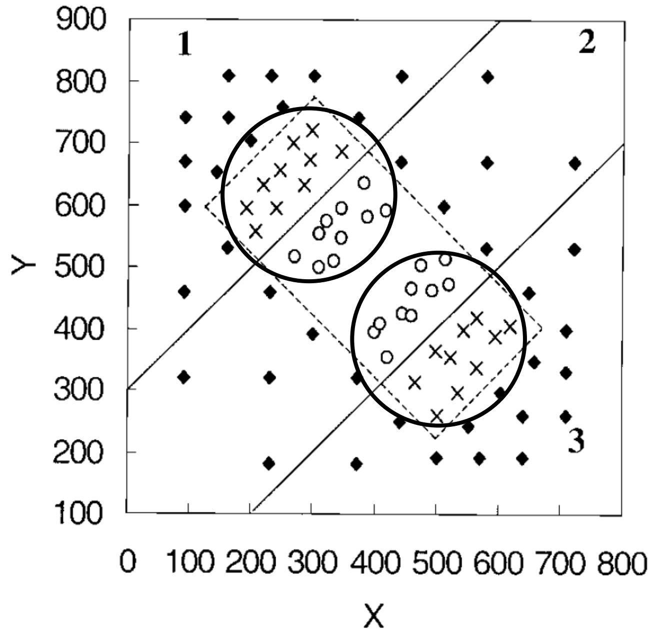


Figure 2: Category space introduced by Yang and Lewandowsky (2003). The solid parallel ascending lines represent partial category boundaries. Training stimuli are enclosed by the dashed box. Category A items are presented as open circles, Category B items as crosses. Training stimuli enclosed by the circles straddling the upper and lower partial boundaries of the space were exclusively presented in the upper and lower contexts, respectively. Transfer stimuli are depicted by filled diamonds. Numerals identify the three diagnostic regions of the stimulus space formed by the partial boundaries. Figure reprinted from Yang, L.-X. and Lewandowsky, S., Context-gated knowledge partitioning in categorization, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 663–679, 2003, published by the American Psychological Association, reprinted with permission.

Perfect classification performance could therefore be achieved by two different strategies. The first strategy, called *context-insensitive* from here on, categorizes stimuli exclusively on the basis of their position in the category space relative to the two partial boundaries. The second strategy, involving *knowledge partitioning*, involves context-dependent application of one partial boundary to the exclusion of the other. The strategies can be differentiated at transfer, when all transfer stimuli (represented as filled diamonds in the figure) are presented once in each context. The context-insensitive strategy places stimuli from Region 2 of the space into Category A and stimuli from Regions 1 and 3 into Category B. In contrast, the knowledge-partitioning strategy classifies stimuli presented in the upper context on the basis of the upper partial boundary, and conversely uses the lower boundary for stimuli presented in the lower context. Hence, knowledge partitioning implies that participants' responses to stimuli from Regions 1 and 3 of the category space should reverse when presented in different contexts.

Experiments using this category structure have consistently found that around a third of participants adopted the knowledge partitioning strategy, whereas another third used the context-insensitive strategy (Lewandowsky et al., 2006; Yang & Lewandowsky, 2003, 2004; see also, Little & Lewandowsky, 2009). Those two groups of participants could not be differentiated on the basis of their training performance (e.g., Yang & Lewandowsky, 2004), implying that the paradigm developed by Lewandowsky, Yang, and colleagues permits acquisition of two different strategies that afford equal performance during training. Moreover, Yang and Lewandowsky (2004) showed by competitive computational modeling with ALCOVE and ATRIUM that these strategies were best characterized by a common set of rules (one instantiating each partial boundary in Figure 2). Thus, the strategies did not differ by relying on different knowledge *per se*, instead, they differed in terms of whether the rules were coordinated on the basis of position x-y space or context. The modular status of these strategies suggests that if restructuring from one to the other is possible, it will likely involve recoordination.

Competing modular response strategies also lead to a key behavioral prediction: That a previously acquired strategy can be preserved (potentially without loss) even after knowledge restructuring to an alternative strategy has occurred. That is, people should be able to revert to using an old strategy even after demonstrably switching to another.

In the present experiments, we adopt the hint-based methodology used by Kalish, Lewandowsky, and colleagues in conjunction with a category structure that supports knowledge partitioning. We test three central predictions about recoordination. First, if people adopt modular response strategies (Yang & Lewandowsky, 2004), knowledge restructuring should conform to a recoordination process. Failure to observe recoordination would imply that the process is unavailable to human learners and that multiple-module models may be more flexible than warranted by data. Second, if recoordination only involves changes in the application of existing knowledge to the task, it follows that the emergence of novel response strategies need not involve additional learning. That is, a hint alone should be sufficient to induce knowledge restructuring; changes in response strategy should arise immediately after a hint, and without requiring additional training. Third, when recoordination drives knowledge restructuring, it should be possible to recover an old strategy by restoring the original coordination state between partial knowledge elements. Like the second prediction, recovery of an old strategy by recoordination should be detectable immediately after a hint, and without requiring re-training.

To foreshadow our principal results and conclusions, we demonstrate that people retain access to old strategies after knowledge restructuring, providing direct empirical evidence of a recoordination process. We also show that people can restructure their knowledge without the need for additional training, suggesting that people may acquire multiple strategies (or the partial knowledge elements needed to instantiate multiple strategies) in parallel, even though only a single strategy spontaneously presents at transfer. Finally, we validate our theoretical claim that recoordination cannot be explained by a single-module architecture (e.g., ALCOVE) by showing that our principal results can only be accommodated by a multiple-module model (ATRIUM).

5. Experiment 1

Experiment 1 examined shifts between knowledge partitioning and context-insensitive response strategies. Based on previous results in this paradigm, we expected a roughly equal number of participants to spontaneously partition their knowledge and to use a context-insensitive strategy, respectively. Halfway through the experiment, a hint was administered that encouraged people to adopt the alternative strategy. Experiment 1 therefore provided a symmetric examination of knowledge restructuring between two alternative strategies.

Recoordination was examined by assessing whether knowledge restructuring could occur in the absence of additional training. Recoordination is evident to the extent that hints immediately induce systematic differences in performance. Finally, we examined the strong prediction made by a recoordination account, that people should be able to immediately reinstate an old strategy in response to a hint (i.e., without requiring re-training). Importantly, if recoordination holds, learning about an alternative strategy is predicted to interfere minimally, if at all, with recovery of a previously acquired strategy.

5.1. Method

5.1.1. Participants, Apparatus, and Design

Forty-four undergraduates from the University of Western Australia participated in exchange for course credit or were remunerated at the rate of A$10/hour.

All experiments were controlled by a Windows PC, running a Matlab program designed using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Experiment 1 involved two sessions of approximately 50 minutes duration each. Sessions were separated by a minimum one-hour break (most participants completed sessions on successive days). Each session began with a 6-block training phase, which was followed by two identical transfer tests, completed successively. Between the transfer tests in each session, participants were provided with a hint that encouraged knowledge restructuring.

The strategy revealed by the hint was conditional on the participant's transfer performance in the first test of each session. Thus, people who were identified as using a knowledge partitioning strategy were given a hint to encourage context-insensitive performance. Conversely, people who where identified as using a context-insensitive strategy were given a hint to induce knowledge partitioning. Experiment 1 involved no between-subjects manipulations that were independent of performance.

5.1.2. Category Space and Stimulus Dimensionality

Stimuli in Experiment 1 were concentric circle pairs (see Figure 3) that were physically defined along two continuous dimensions corresponding to the diameters of the inner and outer component circles. On the basis of a multidimensional scaling study and exploratory modeling (see Appendix B for details), we assumed that people's psychological representation of the stimuli was commensurate with their physical dimensionality. In addition to the dimensions defining the diameters of the inner and outer circles, a third dichotomous dimension representing context was instantiated using color (i.e., red and green).

All stimuli were derived from the category space presented in Figure 3. Training stimuli from the region above the solid black boundary (Areas 1 and 2 in the figure) belonged to Category A, whereas those below the boundary (Areas 3 and 4) belonged to Category B. Context was systematically mapped onto regions of the space during training such that the 20 training stimuli that clustered around the left partial boundary were presented in the left context (e.g., in green), whereas the 20 that clustered around the right partial boundary were presented in the right context (e.g., in red). The proportion of training stimuli belonging to Category A in each training cluster was equal to 0.5; hence, context was not predictive of category membership, but identified the region of the space occupied by a given training item. Perfect training performance could therefore be achieved by one of two equally valid strategies. Performance could be characterized by what we call the *context-insensitive* (CI) strategy, whereby responding is consistent with context-insensitive applica-

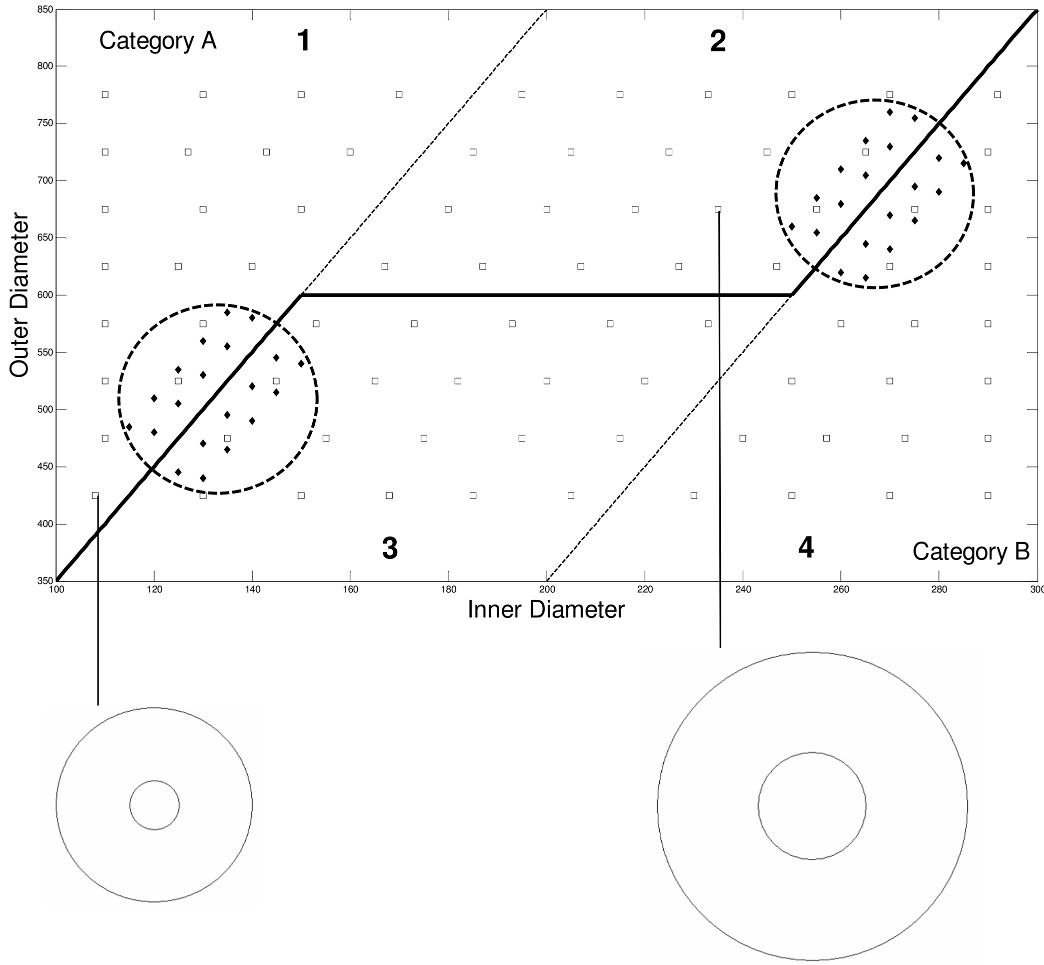


Figure 3: Category space and stimuli used in Experiment 1. The unbroken black line describes the continuous category boundary and regions of the space associated with categories A and B. Dashed lines denote extrapolated partial boundaries expected under knowledge partitioning that divide the space into four diagnostic areas. Filled diamonds represent training stimuli. Training stimuli presented in the left and right training contexts are housed within the left and right circles, respectively. Transfer stimuli are presented as unfilled squares. Two sample stimuli are presented below the category space to illustrate the nature of the task.

tion of the continuous boundary, denoted by the solid black line in Figure 3. Alternatively, people could learn a context-dependent *knowledge partitioning* (KP) strategy, and associate different contexts with different partial category boundaries. Under knowledge partitioning, the left partial boundary would be used to categorize stimuli presented in the left context, and the right partial boundary for stimuli presented in the right context. Although each strategy mandated identical responses during training, they generate distinct response profiles on the transfer test. During transfer, participants were presented with stimuli spanning the entire category space (represented as unfilled squares in Figure 3). Twenty transfer stimuli were drawn from each of the four diagnostic areas, and were presented once in each context, yielding a total of 160 transfer stimuli in each transfer test. For ease of exposition, we refer to the transfer tests in the order in which they were completed. Thus, the transfer tests in the first phase are referred to as Tests 1 and 2, and the tests in the second phase as Tests 3 and 4, respectively.

Application of the continuous boundary strategy would, regardless of context, result in a high probability of responding "A", P(A), for transfer stimuli from Areas 1 and 2, and a low P(A) for stimuli from Areas 3 and 4. Conversely, if people apply the knowledge-partitioning strategy, they are expected to rely exclusively on the left partial boundary (and ignore the right partial boundary) when classifying transfer stimuli presented in the left context. Vice versa, for transfer stimuli presented in the right context; people who partition are expected to rely solely on the right partial boundary to the exclusion of the left partial boundary. Use of a knowledge partitioning strategy is therefore revealed if responses for transfer stimuli in Areas 2 and 3 systematically vary with context such that P(A) for transfer stimuli from Areas 2 and 3 is low when presented in the left context, but high when presented in the right context. P(A) is not expected to vary with context for transfer stimuli from Areas 1 and 4 as both partial boundaries mandate identical responses.

5.1.3. Procedure

Participants were tested individually in a quiet booth. Each trial began with presentation of the word "Ready" for 1000 ms. This was followed by the stimulus, which remained visible until a category response was made via keypress. During training, participants were provided with corrective feedback (the word "Correct" or "Incorrect" presented for 1000 ms) immediately following each response. Feedback was withheld during transfer. Trials were separated by a 1000 ms blank interval. Participants were provided with self-paced breaks after every block of 40 trials. Category responses were recorded using the F and J keys. Both color-to-context and category-to-key mappings were counterbalanced across participants.

Within each training block, all 40 training stimuli were presented once in a random order. All stimuli from the right training cluster were presented solely in the right context, while stimuli from the left training cluster were presented solely in the left context. Each transfer test involved two presentations of each transfer stimulus, in a random order, once in each context, resulting in 160 transfer trials. In each session, after the first transfer test, participants were provided with a hint to encourage knowledge restructuring. Performance in the first transfer test determined the content of the hint, and was assessed online by computing the Euclidean distance between a participant's response profile and two pre-defined centroids corresponding to the KP and CI strategies, respectively. The hint revealed the contrast strategy; that is, the strategy that was most distant from that person's response profile.

People with response profiles furthest from the CI strategy were encouraged to adopt the CI strategy, and were informed that context (color) was not relevant to category decisions and should not be incorporated into their decision strategy, and that category membership was determined solely by the relative sizing of the inner and outer circles. Conversely, people whose response profile was furthest from the KP strategy were told that categorization could be achieved by examining the relative sizing of the inner and outer stimulus dimensions, but that the relevant size relation differed according to context (color). If a participant's profile was equidistant to the KP and CI options, the hint was determined randomly (only one person retained for analysis fell into this category). Participants were given a few minutes to study the hint, which was presented to them in writing. To ensure comprehension, participants were required to summarize the revealed strategy without referring to the written statement. Once an accurate summary had been produced, participants were encouraged to adopt the revealed strategy and continued the experiment.

5.2. Results and Discussion

5.2.1. Data Screening

To be included for analysis, a participant's Session 1 accuracy in the training block immediately preceding the first transfer test (i.e., training block six) had to significantly exceed chance (approximately 0.65 assuming binomially distributed responses for the $n = 40$ training items, with $p = .5$, and $\alpha = .05$). The analysis retained 29 participants. Although this reduced sample size may raise concerns regarding the generality of the results, selective analysis of learners has ample precedent in the literature (e.g., Erickson & Kruschke, 1998; Johansen & Palmeri, 2002; Medin & Schwanenflugel, 1981; Nosofsky, Clark, & Shin, 1989). None of our conclusions are altered if inclusion criteria are relaxed (e.g., by retaining any participant whose training performance exceeded 0.5). Later, we replicate the results of Experiment 1 using a different category structure, which promoted near-ceiling learning performance.

5.2.2. Initial Training Performance

During the first session, training performance increased from .61 in the first block to .76 in the final block. A within-subjects ANOVA with block as a factor returned a significant main effect, $F(5, 140) = 13.01$, $MS_e = .01$, $p < .001$, $\eta_p^2 = .32$. In the final block of training in the first session, there was no significant difference between participants who were later identified as applying the KP strategy ($M = .77$) versus those who used the CI strategy ($M = .75$), $t(22) = .52$, $p = .61$, $r^2 = .01$ (cf. Yang & Lewandowsky, 2004). Training performance throughout the second session was comparable to the final level observed at the first session ($M = .75$ across blocks) and is not presented in detail.

5.2.3. Individual Differences during Transfer Test 1

Preliminary examination of individual response profiles revealed considerable variation in strategy use. To identify the principal response strategies, we performed a k-means cluster analysis using each participant's transfer responses as input (by recoding "A" and "B" responses into 1s and 0s, respectively). The clustering algorithm was seeded with three pre-defined starting points corresponding, respectively, to context-dependent use of the partial category boundaries (KP strategy), context-insensitive use of the continuous category boundary (CI strategy), and chance performance on all items. Using a Euclidean distance measure, the clustering algorithm computed final centroid positions that maximized within-cluster similarity and between-cluster differences. Examination of the final three cluster centroids indicated that one group of people utilized a context-dependent strategy, consistent with what would be expected under knowledge partitioning (called the Initial-KP group from here on; $N = 10$); another that responded primarily on the basis of the continuous category boundary (Initial-CI group; $N = 17$); and a final group whose response profile was unrecognizable ($N = 2$).[4] The data from the latter group were not considered any further. Correspondence between the iterated clustering algorithm and the online measure of strategy use was very high; identical strategy classifications were returned for 24 of the 27 remaining participants. For consistency, only those participants who were identically classified by both the online measure and cluster analysis were

[4] Initializing the clustering algorithm with three random starting points yielded nearly identical cluster allocations.

retained. Figure 4 displays item-wise performance in the first transfer test for participants in the Initial-KP and Initial-CI groups (Ns = 8 and 16, respectively). The figure shows that the Initial-CI group performed similarly between contexts, whereas the Initial-KP group relied on one or the other partial boundary, as determined by context, to classify all items.

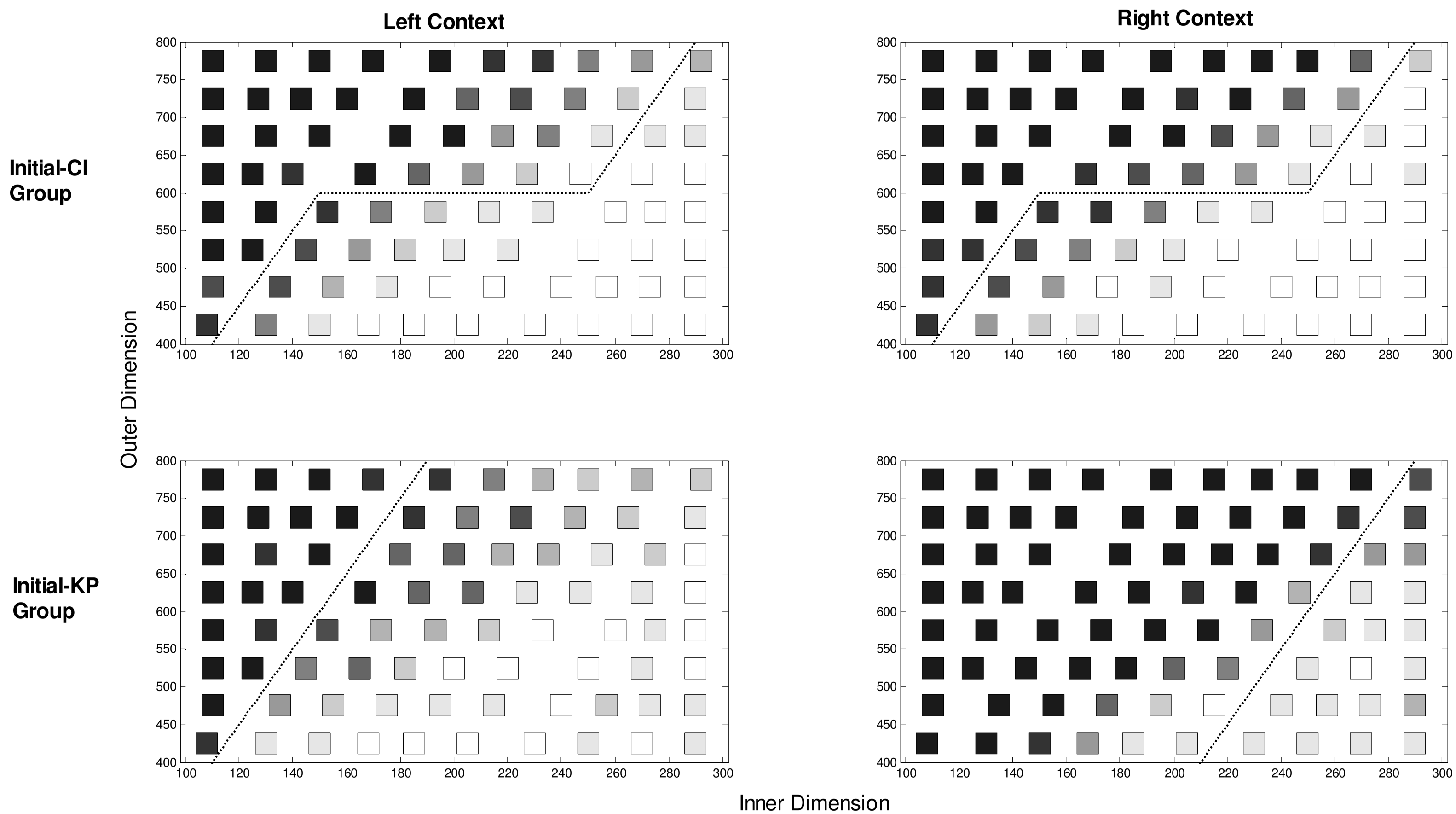


Figure 4: Item-wise P(A) in each context in the first transfer test for the Initial-KP and Initial-CI groups in Experiment 1. Strategy-appropriate category boundaries are included for emphasis. Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

We statistically confirmed the pattern in Figure 4 by first averaging P(A) across transfer stimuli within each diagnostic area for each context, yielding a mean P(A) for each diagnostic area in each context. These values were then averaged across participants to obtain group-wise means. A Group (Initial-KP vs. Initial-CI) $\times$ Context $\times$ Area between-within ANOVA returned a significant three-way interaction, $F(3, 66) = 16.64$, $MS_e = .01$, $p < .001$, $\eta_p^2 = .43$, reflecting the differential between-group sensitivity to context. To confirm that the two groups performed differently within each context, we conducted two follow-up Group $\times$ Area between-within ANOVAs, one for each context. Significant interactions were observed in the left, $F(3, 66) = 6.13$, $MS_e = .02$, $p < .001$, $\eta_p^2 = .22$, and the right contexts, $F(3, 66) = 9.88$, $MS_e = .01$, $p < .001$, $\eta_p^2 = .31$. We therefore conclude that the Initial-KP and Initial-CI groups adopted different response strategies: the KP and CI strategies, respectively.

5.2.4. Knowledge Restructuring & Strategy Recovery

To simplify further analysis of strategy use across transfer tests, we introduce a measure of *context sensitivity* that focuses exclusively on stimuli from Areas 2 and 3. These are stimuli for which the KP strategy predicts responding to change with context, and thus serves as a proxy measure for use of the KP strategy. For these stimuli, we computed the average item-wise difference in P(A) between contexts. On the first transfer test, people in the Initial-KP group showed a high level of context sensitivity ($M = .54$), whereas those from the Initial-CI group showed a lower level of context sensitivity ($M = .06$). Figure 5 shows context sensitivity across all transfer tests for both groups. It is clear from the figure that the patterns of context sensitivity between the Initial-KP and Initial-CI groups differed considerably, implying that the hints were effective at eliciting knowledge restructuring, regardless of which strategy people initially applied.

We performed two sets of diagnostic tests for recoordination. Recall that recoordination, unlike relearning, predicts that knowledge restructuring can occur immediately after a hint, and without requiring additional training. Thus, recoordination predicts performance differences between the two successively presented transfer tests (Tests 1 and 2), despite the absence of intervening training. Figure 5 shows that revelation of an alternative strategy had a clear, immediate, and symmetrical effect on context sensitivity. In support, in the Initial-KP group, context sensitivity between Tests 1 and 2 significantly decreased (M = -.43), reflecting knowledge restructur-

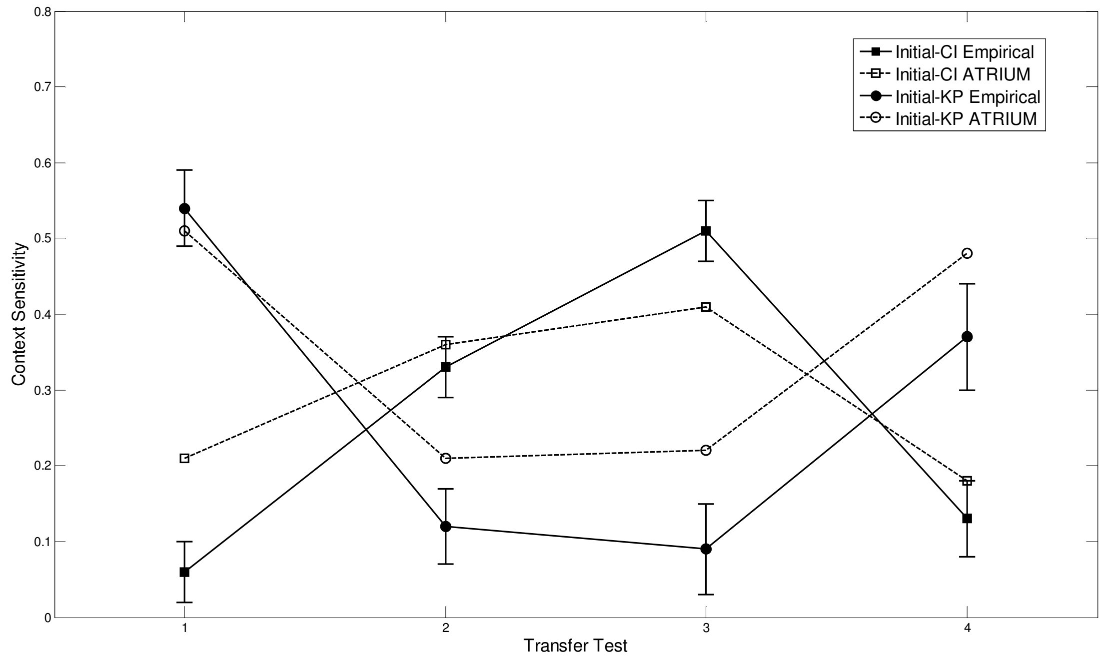


Figure 5: Observed and predicted levels of context sensitivity across transfer tests in Experiment 1. Solid black lines display data, dashed lines display ATRIUM's predictions.

ing from the KP to the CI strategy, $t(7)$ =-5.35,p = .001, $r^2 = .80$. By contrast, in the Initial-CI group, there was a significant increase ($M = .26$) in context sensitivity as people restructured from the CI to the KP strategy, $t(15) = 5.93, p < .001, r^2 = .70$. This is precisely the pattern predicted by recoordination, and suggests that either the two strategies were simultaneously available to participants after the initial training (e.g., as different representational structures people toggled between), or alternatively, that the strategies were reliant on common elements of partial knowledge, which were coordinated in different ways following the hint.

Considering the effects of additional training, Figure 5 shows a further increase in context sensitivity between Tests 2 and 3 for the Initial-CI group. For the Initial-KP group, by contrast, there was little further change in context sensitivity beyond that observed immediately following revelation of the CI strategy. This is not surprising, as the Initial-KP group was already minimally sensitive to context after the hint. The fact that additional training enhanced the effects of the initial restructuring (at least in the Initial-CI group) is consistent with the notion that people refined the strategies revealed by the hint during the training period in Session 2.

Our second critical test for recoordination examined changes in context sensitivity between Tests 3 and 4. Recall that participants were provided with another hint in Session 2, directing them away from the strategy used on Test 3, and hence toward their original strategy (i.e., that used in Test 1). Recoordination predicts robust strategy recovery, because all that is required for knowledge restructuring is recovery of the original coordination state of partial knowledge elements. Relearning (at best) predicts poor strategy recovery, as acquisition of the second strategy might interfere with recovery of the first (e.g., Macho, 1997). Consistent with recoordination, context sensitivity in the Initial-KP group significantly increased between Tests 3 and 4 (M = .28), indicating restructuring from the CI to the KP strategy, $t(7) = 2.93, p = .02, r^2 = .55$. Analogously, context sensitivity in the Initial-CI group significantly decreased between Tests 3 and 4 (M = -.38), reflecting restructuring from the KP strategy, and recovery of the CI strategy, $t(15) =$ -6.77, $p < .001, r^2 = .75$. This pattern indicates renewed restructuring in the direction of people's original response strategy, implying retention of the strategy despite having previously restructured to an alternative, and after having completed 240 training trials using that alternative. To test whether the restructuring observed between Tests 3 and 4 involved recovery of people's original strategies, we computed item-wise correlations between performance on Tests 1 and 4 for both groups[5]; the correlations were quite high, r = .81 (Initial-KP group), and r = .95 (Initial-CI group).

5.2.5. Summary of Results

Experiment 1 showed, for the first time, that people can fluidly restructure their knowledge, shifting between context-dependent (KP) and context-insensitive (CI) response strategies without needing additional training. The results are well described by a recoordination process but are at odds with

[5]At the level of individual items, the KP and CI strategies make identical predictions for 75% of transfer stimuli. Because this high degree of overlap would artificially inflate correlations, the only responses considered by this and all similar subsequent correlational analyses were those associated with stimuli for which the strategies make divergent predictions (i.e., stimuli from Area 2 presented in the left context and those from Area 3 presented in the right context).

a relearning process. Our results build on existing studies, showing that knowledge restructuring extends beyond situations involving strategies that differ in terms of the maximum level of accuracy they afford (cf. Johansen & Palmeri, 2002; Kalish et al., 2005; Lewandowsky et al., 2000); in Experiment 1, the KP and CI strategies both permitted perfect accuracy during training.

The fact that the observed restructuring was symmetrical (i.e., people could shift from the KP to the CI strategy as well as in the reverse direction) demonstrates that knowledge restructuring is not constrained by initial strategy use. This result implies that people may be less reluctant to abandon simple strategies than has been previously thought Kalish et al. (2005); Lewandowsky et al. (2000). In our experiment, the CI strategy can be considered the *de facto* complex strategy, as knowledge partitioning is usually viewed as a means of simplifying a complex task (e.g., Kalish et al., 2004; Lewandowsky et al., 2002; Little & Lewandowsky, 2009), and yet people readily restructured both to and from the KP strategy. That people can restructure even after initially partitioning their knowledge is also perhaps counterintuitive given that previous work has shown that knowledge partitioning can persist even in expertise (Lewandowsky & Kirsner, 2000).

Symmetrical knowledge restructuring is consistent with either the parallel acquisition of the KP and CI strategies, or acquisition of a set of partial knowledge elements that could be used to construct either strategy. The latter implies that people applying different strategies possess a common underlying set of partial knowledge, and that their behavior differs only by virtue of how partial knowledge is coordinated and applied to the task. We prefer the latter explanation, as it accords closely with the computational modeling results reported later.

Although Experiment 1 supported the notion of a recoordination process, there are two important limitations that must be addressed. First, the generality of the result is potentially limited by the relatively small number of subjects who were included in the final analysis. The task was deliberately engineered to be quite difficult (i.e., to generate sufficient error) in order to maximize the likelihood of observing knowledge restructuring (Kalish et al., 2005). An unintended consequence was that many participants failed to reach the learning criterion. It is therefore unclear whether recoordination depends on factors such as error-prone performance or task difficulty. Second, the nature of the CI strategy introduced some ambiguity as to whether people explicitly represented the continuous boundary, or if they only represented the component partial boundaries and selectively applied them without regard to context. Although we prefer the latter interpretation on the strength of related modeling work with ATRIUM (Yang & Lewandowsky, 2004), we acknowledge that the category space of Experiment 1 does not guarantee that people were employing the CI and KP strategies in the exact manner in which we described them.

Experiment 2 addressed these concerns by using a modified category space involving simple unidimensional boundaries. The use of unidimensional instead of multidimensional boundaries should simplify the task insofar that people are initially biased toward lower-dimensional solutions to categorization problems, as posited by several models (e.g., Love et al., 2004; Nosofsky et al., 1994). We provided people with quite detailed hints describing different rule-based strategies before commencing the task, obviating the need for an individual differences analysis to determine strategy use. That is, unlike in Experiment 1, we exercised experimental control over what strategy people initially acquired. If the modular structure of the strategies was the driving factor behind the knowledge restructuring observed in Experiment 1, recoordination should be observed even when performance is highly accurate and category boundaries are unidimensional. By contrast, if recoordination only emerges in quite difficult tasks or when performance is error-prone, Experiment 2 would be expected to produce very little evidence of recoordination.

6. Experiment 2

Experiment 2 utilized a novel category space based on that of Aha and Goldstone (1992), and did not involve a continuous boundary, ensuring unambiguous identification of the CI strategy. Moreover, the partial boundaries characterizing the space used in Experiment 2 were incommensurate, in that they predicted conflicting category responses to stimuli in the top left, and bottom right regions of the space (see Figure 6). Because the boundaries could not be incorporated into an integrated coherent strategy, the space required a modular strategy, regardless of the content of that strategy (i.e., KP or context-insensitive).

6.1. Method

6.1.1. Participants and Design

Forty-eight undergraduates from the University of Western Australia participated in exchange for course credit or were remunerated at a rate of A$10

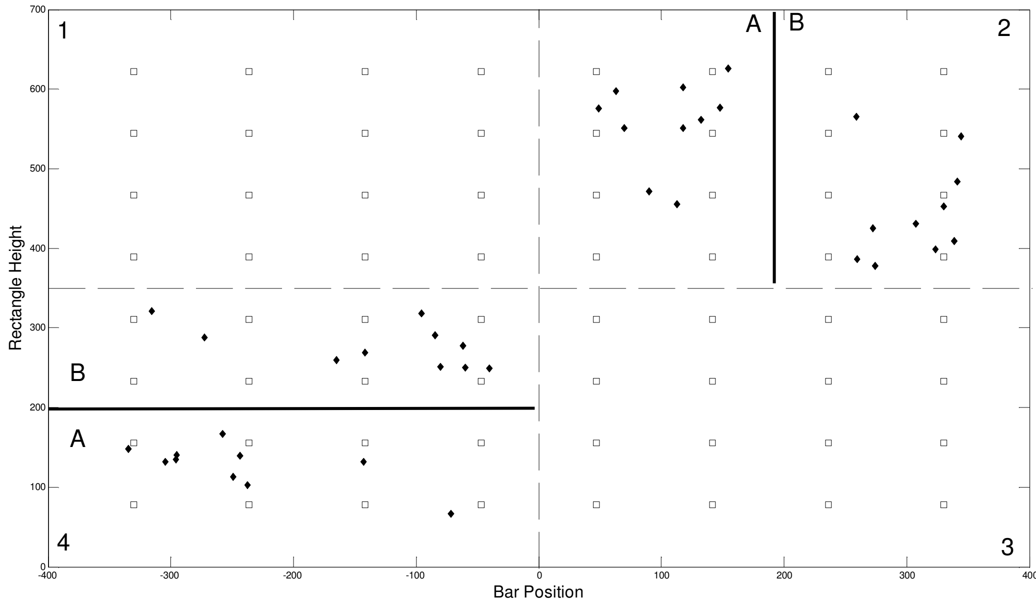


Figure 6: Category space used in Experiment 2. Filled diamonds denote training stimuli, open squares denote transfer stimuli. Solid lines are the partial rule boundaries. Dashed lines divide the space into four diagnostic quadrants, which are numbered in the figure.

per hour.

The structure of Experiment 2 was similar to that of Experiment 1; the experiment comprised two 1-hour sessions, each involving a training period followed by two successive transfer tests. Unlike Experiment 1, where hints were determined by performance on an initial transfer test, the hints in Experiment 2 were fixed depending on condition. At the outset of the first session, participants were provided with a hint describing a context-dependent KP strategy (*KP-first* condition), or a context-insensitive strategy (*CI-first* condition). After the first transfer test in each session, the alternative strategy was revealed (i.e., the CI strategy was revealed in the KP-first condition; the KP strategy was revealed in the CI-first condition). As in Experiment 1, the hint describing the alternative strategy was reinstated at the start of Session 2, and participants were then able to practice that strategy with feedback. Partway through the second session, the hint presented at the outset of the experiment (i.e., the original hint from the first session) was provided.

6.1.2. Stimuli, Category Space, Strategies, and Hints

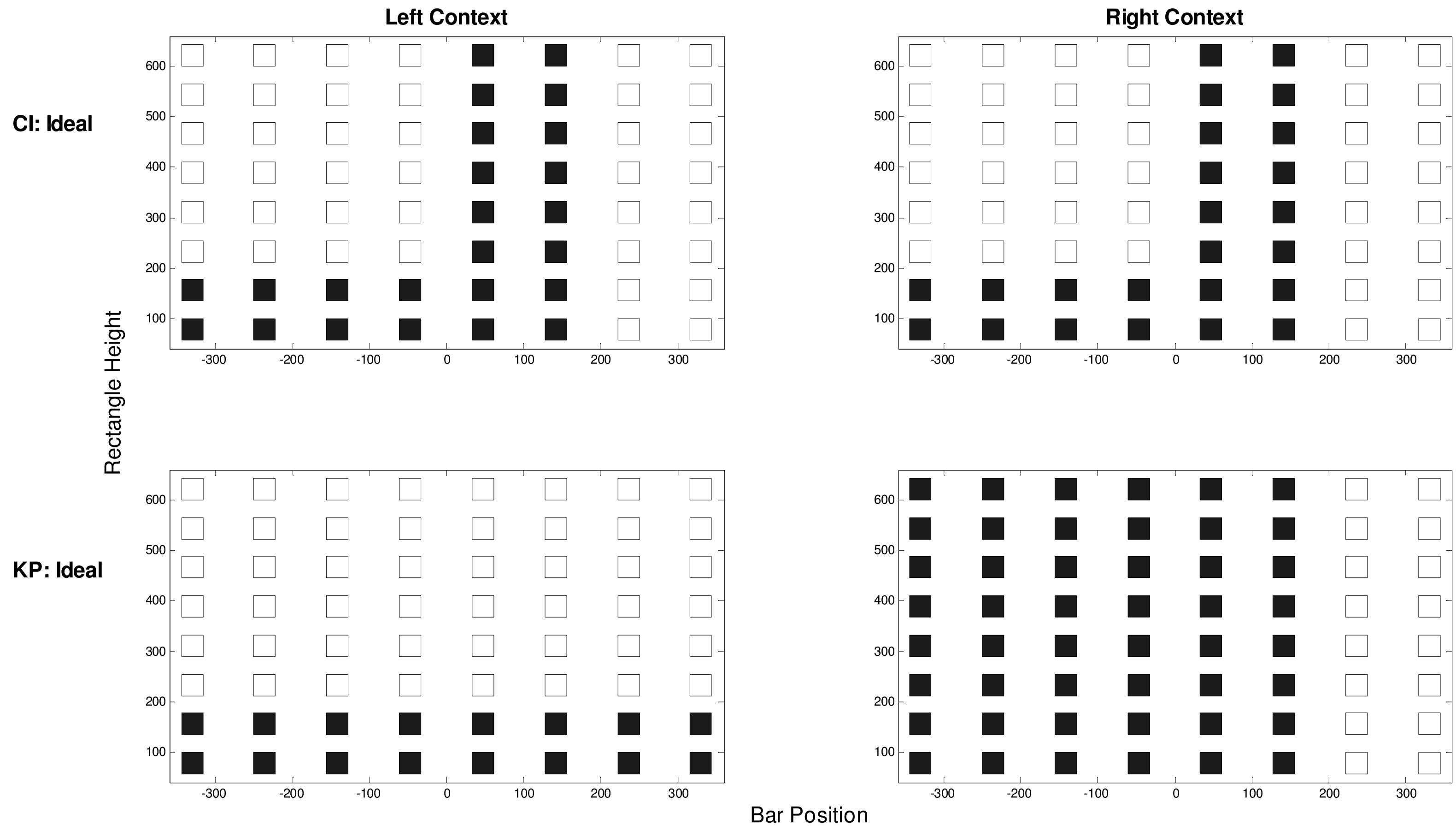


Figure 7: Ideal response profiles associated with the context-insensitive (top row) and knowledge partitioning (bottom row) strategies. Performance in the left and right contexts are shown in the left and right columns of panels, respectively.

Stimuli used in Experiment 2 were rectangles that could vary along two dimensions: The height of the rectangle, and the horizontal positioning of a small bar offset along the base. These stimuli have been used extensively in category learning tasks, and are known to be perceptually separable (Erickson & Kruschke, 1998; Kruschke, 1996; Lewandowsky et al., 2006).

The category space used in Experiment 2 is presented in Figure 6. It comprises two unidimensional boundaries. The left boundary is orthogonal to the vertical axis (rectangle height), whereas the right boundary is orthogonal to the horizontal axis (bar offset position). Importantly, Category A stimuli are placed below the left boundary, whereas they are placed to the left of the right boundary. This ensures that the two boundaries cannot be integrated in a coherent way. Like the space used in Experiment 1, subsets of training stimuli (40 in total; filled diamonds in Figure 6) straddled each partial boundary, and a binary context dimension was mapped onto the subsets. Although both stimulus dimensions are thus equally relevant to the task, different dimensions were relevant to different subsets of stimuli. The space therefore permitted strategies analogous to the KP and CI strategies from Experiment 1. Idealized generalization profiles for the two strategies are shown in Figure 7. In this experiment, the KP strategy is characterized by context-dependent use of the horizontally and vertically oriented partial category boundaries. In the left context, performance is dependent on the left rule (i.e., the rule along the y, rectangle height, dimension). In the right context, performance is determined by the right rule (i.e., the rule along the x, bar position, dimension). By contrast, the CI strategy predicts that rule use will vary according to the location of the stimulus along the x dimension (as specified in the hint, detailed below). For low values on x (i.e., bar position on the left-hand side of the stimulus), the left rule applies. For high values on x (i.e., bar position on the right-hand side of the stimulus), the right rule applies. Because the CI strategy holds that coordination of the two rules is context-invariant, a signature "backwards-L" response profile should obtain regardless of context under the CI strategy.

To facilitate model-based analysis, the category space and intended response strategies were deliberately designed to pose challenges to a single-module exemplar-based model like ALCOVE. Because ALCOVE has no mechanism for stimulus-dependent attentional weightings (though see Rodrigues & Murre, 2007), it must attend to both x and y dimensions regardless of whether it is implementing the KP or CI strategy. This prevents ALCOVE from extrapolating the vertical "A" component of the CI strategy (see top panels of Figure 7), whilst also preventing successful extrapolation of that boundary when using the KP strategy in the right context — ALCOVE assigns stimuli with mid-level values on the y dimension to Category B due to their proximity to the B items from the left training cluster. By contrast, ATRIUM can very naturally produce the idealized response profiles via coordination of its rule modules.

The hints given to participants at the start of the experiment were analogous to the KP and CI strategies from Experiment 1. The initial hint was dependent on condition, which was randomly assigned. Both hints stated that for any given stimulus, only a single dimension determined category membership. However, the relevant dimension depended on either stimulus color (KP strategy), or whether the bar position was on the left or right half of the stimulus (CI strategy). For example, the hint for the KP strategy stated that for (say) red stimuli, rectangle height exclusively determined category membership, but for (say) green stimuli, bar position exclusively determined category membership. Similarly, the CI strategy hint stated that if bar position was on the left-hand side of the stimulus, rectangle height determined category membership, but if the bar was on the right-hand side, bar position uniquely determined category membership. Neither hint specified any particular "cut-off" value for either dimension, requiring participants to learn the locations of the partial boundaries through trial and error.

6.1.3. Procedure

The procedure was similar to that of Experiment 1, with the following exceptions. (1) Hints were administered at the outset and assignment to initial hint condition was random; 25 participants were assigned to the KP-first condition, 23 to the CI-first condition. (2) We permitted an "early exit" from the first session of training; training was aborted if, after completion of at least 4 full training blocks, the participant made 40 consecutive correct responses. (3) The number of training blocks in the second session was reduced from 6 to 2 (no early exit was available from the second session). (4) Because this experiment was embedded within a larger project involving collection of individual differences data, participants were randomly assigned to 1 of 8 pre-loaded (but randomly permuted) training sequences.

6.2. Results

6.2.1. Data Screening

The accuracy criterion used in Experiment 1 retained 47 of the 48 participants. The high number of participants retained for analysis reflects the relative ease of the task used in Experiment 2.

6.2.2. Training Performance

A large number of participants achieved the criterion for early exit from training (18 from the KP-first, 13 from the CI-first condition). This reflects highly accurate overall performance. Of the participants who did not exit early from training, mean performance in the final training block of session 1 was very high, $M = .90$ in the KP-first condition, $M = .91$ in the CI-first condition. The poorest performers in each condition were still quite accurate, .70 and .83 final block accuracy for the KP-first and CI-first conditions, respectively. Accuracy in the final block of the session 2 training was comparably close to ceiling (Ms = .95 and .94 for the KP-first and CI-first conditions, respectively.). Clearly, the task was learned very well. It follows that knowledge restructuring observed in the current experiment cannot be explained by performance error. Further, observing recoordination would highlight the voluntary element of this form of knowledge restructuring.

6.2.3. Strategy Differences in Test Performance

Because strategy use was controlled experimentally (via hint condition), an individual difference analysis was unnecessary[6]. Instead, our analysis of transfer test performance focuses on differences in strategy use between the KP-first and CI-first conditions. Performance averaged across participants in each condition in the first transfer test are presented in Figure 8. It is clear that participants in the CI-first condition responded using the CI strategy, whereas those in the KP-first condition relied on knowledge partitioning (cf. Figures 7 and 8).

[6]There was one individual difference issue which impacted the modeling of the Experiment 2 data, but did not affect interpretation of strategy use: In the CI-first group, in the right context of Test 2, people positioned the vertically oriented boundary in one of two positions along the x dimension. We discuss this issue in detail in the Computational Modeling section of the paper.

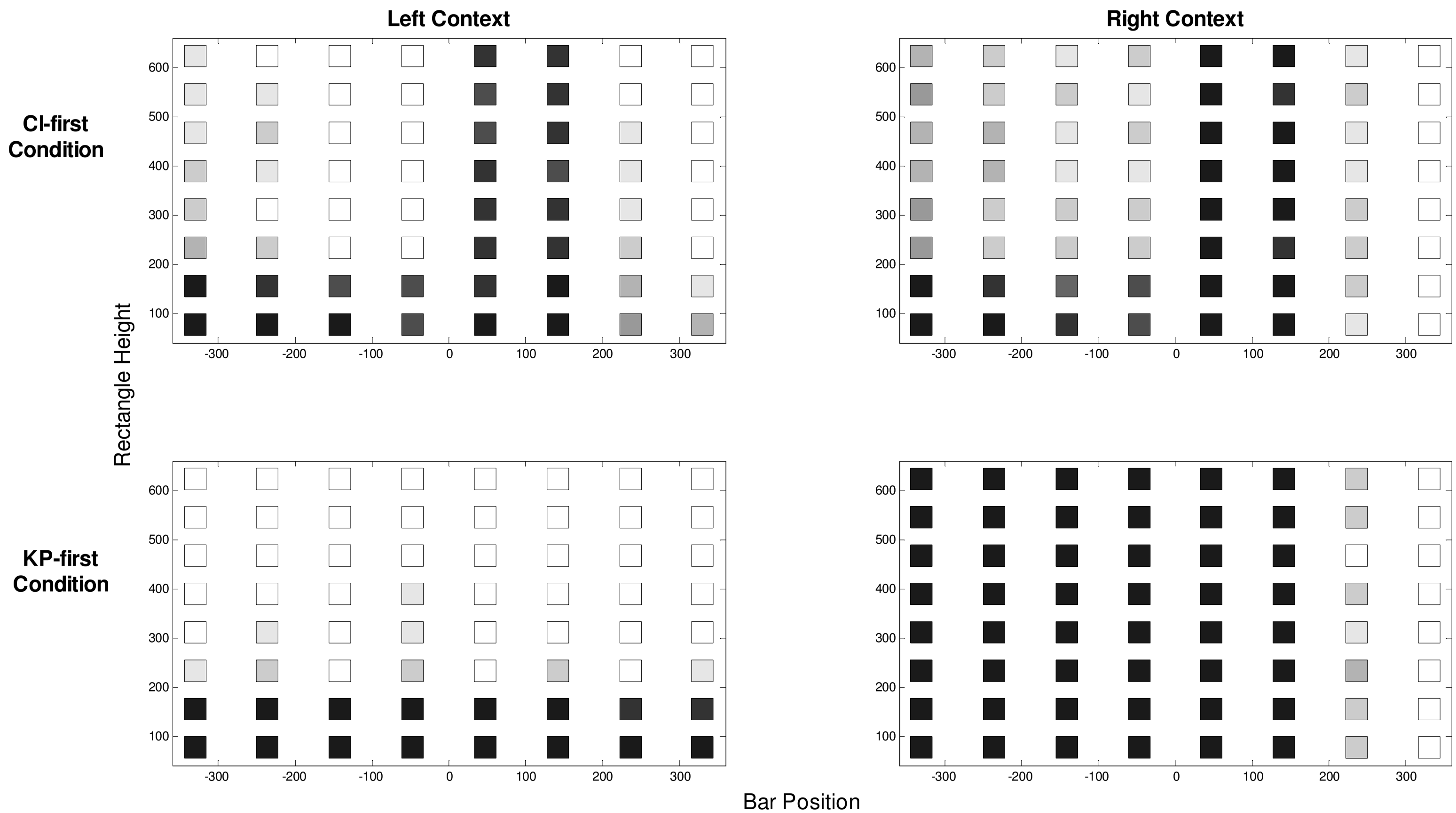


Figure 8: Item-wise P(A) in each context in the first transfer test for the KP-first and CI-first conditions in Experiment 2. Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

We statistically confirmed the between-condition differences in categorization strategies via an analysis similar to that carried out for Experiment 1. We divided the space into four diagnostic regions by aggregating, for each context, P(A) across transfer stimuli in the four quadrants of the space (i.e., the regions defined by the dashed lines and labeled numerically in Figure 6). We then averaged across participants. A 2 (Condition) $\times$ 2 (Context) $\times$ 4 (Quadrant) between-within ANOVA returned a significant 3-way interaction, F (3, 135) = 42.52, MS_e = .01, p < .001, η_p^2 = .49. Follow-up Condition $\times$ Quadrant between-within ANOVAs on performance within each context revealed condition-specific patterns of responding; significant interactions were observed in the left context, $F(3,\ 135) = 38.08$, $MS_e = .01$, $p < .001$, $\eta_p^2 =$ .46, and the right context, $F(3,\ 135) = 40.08$, $MS_e = .03$, $p < .001$, $\eta_p^2 =$ .47. The pattern of results demonstrates that the hints were effective at determining the manifest categorization strategy in the first transfer test. Our subsequent analyses focus on the measure of context sensitivity introduced in Experiment 1.

6.2.4. Knowledge Restructuring

Recall that our context sensitivity measure focuses only on stimuli that would be categorized differently across contexts if the KP strategy were perfectly applied. Figure 9 tracks the changes in context sensitivity across the four transfer tests for each condition. It is clear that participants were not only able to immediately restructure their knowledge and apply the contrast strategy in response to the hint, but that subsequent supervised learning of the contrast strategy did not interfere with reinstatement of their original strategy. We again performed two sets of diagnostic tests of recoordination by examining changes in context sensitivity between Tests 1 and 2 (test for initial knowledge restructuring) and Tests 3 and 4 (test for strategy recovery) for the KP-first and CI-first conditions.

Like in Experiment 1, the hints had a clear, immediate, and symmetrical effect on performance. In the KP-first condition, there was a clear drop in context sensitivity between Tests 1 and 2 (M = -.89), reflecting knowledge restructuring from the KP to the CI strategy, $t(24)$ = -42.19,$p < .001, r^2 = .99$. By contrast, in the CI-first condition, context sensitivity increased between Tests 1 and 2 (M = .48), reflecting restructuring from the CI to the KP strategy, $t(21) = 5.69, p < .001, r^2 = .61$. Analogous comparisons of context sensitivity in Tests 3 and 4 were also consistent with strategy recovery mediated by recoordination. In the KP-first condition, context sensitivity between

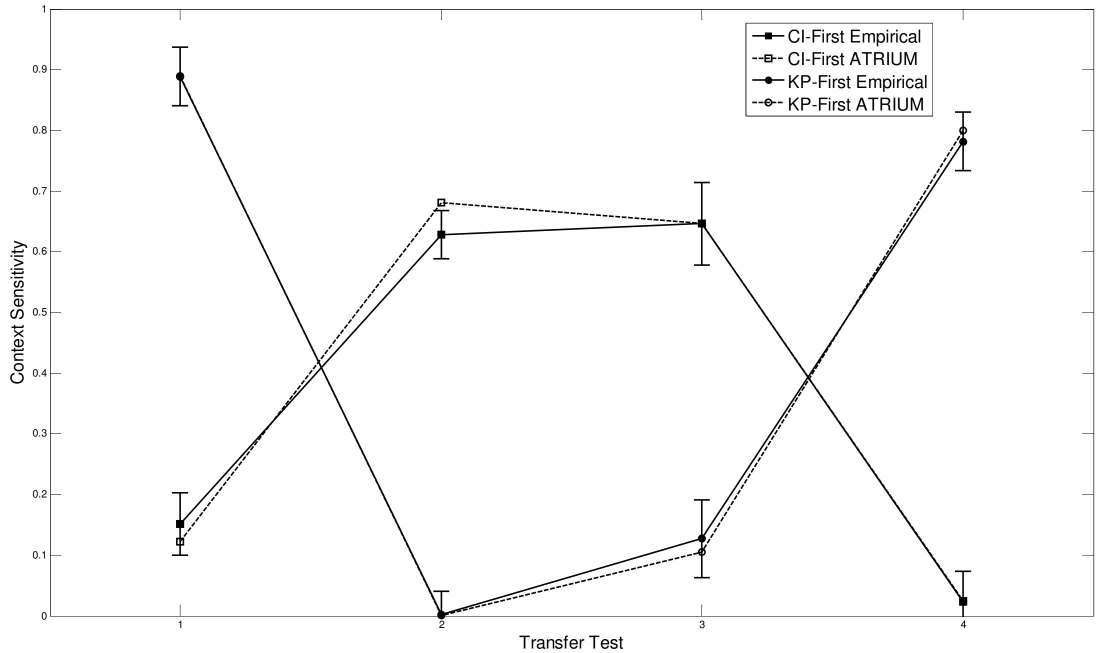


Figure 9: Observed and predicted levels of context sensitivity across transfer tests in Experiment 2. Solid black lines display data, dashed lines display ATRIUM's predictions.

Tests 3 and 4 increased (M = .66), as people reinstated the KP strategy, $t(24) = 6.31, p < .001, r^2 = .62$. In the CI-first condition, there was a reduction in context sensitivity (M = -.62), as people reverted back to the CI strategy, and away from knowledge partitioning, $t(21)$ =-7.14,$p < .001, r^2 = .71$. To determine whether the knowledge restructuring between Tests 3 and 4 involved recovery of people's original response strategy, we computed correlations analogous to those from Experiment 1 for the KP-first and CI-first conditions. In both cases, the correlations were very high, r = .99 (KP-first), r = .96 (CI-first), implying strategy recovery.

6.2.5. Summary of Results

Experiment 2 replicated and extended the recoordination observed in Experiment 1 to a novel set of stimuli, a new category space, and a larger sample. Experiment 2 also showed that recoordination is not restricted to situations involving error-prone task performance; training performance was clearly at ceiling. People fluidly restructured their knowledge in response to the hint, shifting response strategies even in the absence of additional training. This is strongly suggestive of the co-acquisition of the KP and CI strategies during the initial training phase. Interestingly, Experiment 2 showed that the contrast strategy (or the representational elements necessary for implementing the contrast strategy) was acquired despite explicit instruction to learn a single specific strategy during initial training.

By constraining the order of strategy acquisition via instruction, the results are also quite surprising in that there is no evidence for either proactive or retroactive interference of one strategy on another. This is particularly striking when considered against the substantial backdrop of previous research that has identified both proactive and retroactive interference effects in situations involving sequential learning (e.g., Hartnett & Gelman, 1998; Kruschke, 1996; Macho, 1997; Vosniadou & Brewer, 1994). Implementing one strategy did not appear to hinder people's capacity to use an alternative strategy regardless of whether they were shifting to a novel strategy (between Tests 1 and 2) or recovering a previously encountered strategy (between Tests 3 and 4) — precisely as predicted on the basis of recoordination.

The extent to which the behavioral data from Experiments 1 and 2 support a recoordination account of knowledge restructuring rests on two key assumptions about the KP and CI response strategies. First, on the basis of simulation results from similar experiments, we have assumed that all strategies are structurally modular (Yang & Lewandowsky, 2004). Second, we have assumed that the observed knowledge restructuring was driven by a mechanism that controls the coordination of partial knowledge elements (e.g., ATRIUM's gating mechanism). The empirical data alone cannot confirm either of these assumptions; rather, they must be assessed via computational modeling.

7. Computational Modeling

Our principal empirical finding is the fluid knowledge restructuring between mutually incompatible categorization strategies, coupled with retention and subsequent recovery of an old response strategy, even after successfully learning and implementing an alternative. We now extend the simulations reported in the Introduction by reporting comparative computational modeling of the transfer data from Experiments 1 and 2 using ALCOVE (Kruschke, 1992) and ATRIUM (Erickson & Kruschke, 1998, 2002a). Our model-based analysis addresses four key predictions made by recoordination. First, recall our emphasis on the link between recoordination and modular model architecture raised in the Introduction. If the data are best explained by a recoordination process, the data should be better fit by a multiple-, rather than a single-module model. Hence, ATRIUM is predicted to provide a better account of the data. Second, ATRIUM's account of knowledge restructuring is predicted to be based on recoordination of its modules; specifically, by applying its rule modules to the task in different, strategy-dependent ways. Thus, knowledge restructuring is expected to occur at the level of ATRIUM's gating mechanism, not within the associative weight matrices of its component modules. Third, because recoordination assumes structurally modular strategies, differences in overt strategy need only reflect differences in the coordination, but not the content, of partial knowledge. Thus, functional differences in task representation need not be present at the level of partial knowledge. This implies that if ATRIUM is fit to data from different groups of people implementing different response strategies (e.g., context-insensitive vs. knowledge partitioning), the pattern of weights used to perform the task (e.g., rule module weights) should be highly correlated across fits despite between-group differences in overt strategy use. The same argument can be applied within groups that implement different strategies at different points in time, leading to a fourth prediction. If common partial knowledge underpins multiple strategies, weights learned to implement an initial strategy should be largely unaffected by subsequent learning about the contrast strategy. Thus, partial knowledge weights from before and after training-based acquisition of an alternative strategy should be highly correlated within a given group of people. Before examining these specific predictions, we provide technical overviews of the two models.

7.1. ALCOVE: An Exemplar Model of Categorization

ALCOVE categorizes stimuli on the basis of their similarity to exemplars stored in memory. Training exemplars are associated with category responses via learned connections. Exemplar activation is driven by similarity to the input such that exemplars most similar to the stimulus achieve the highest activations. If $h_{e_{js}}$ describes the positioning of exemplar node e_j along dimension s in psychological space, activation of exemplar node e_j, given input

i is described by,

$$a_{e_j} = \exp\left[-c\left(\sum_s \alpha_s \left|h_{e_{js}} - i_s\right|\right)\right], \tag{1}$$

where the specificity, c, is a freely estimated parameter. Note that Equation 1 assumes an exponential similarity gradient and a city-block distance metric (the absolute psychological distance along different stimulus dimensions is combined additively to determine the psychological distance between two points). The use of a city-block metric implies psychologically separable stimulus dimensions (Shepard, 1987), and is consistent with the multidimensional scaling results reported in Appendix B as well as previous uses of the stimuli from Experiment 2 (e.g., Erickson & Kruschke, 1998; Kruschke, 1996). The distance between two points in psychological space is modulated by learned dimension-specific attention weights, α_s.

Exemplar nodes are connected to category nodes, one for each possible category response, via learned exemplar-to-category association weights. The overall activation of each category node is the weighted sum of the activation of each exemplar node, modulated by the association weights. The relative activation levels of the category nodes determine ALCOVE's response, which is expressed as the probability of an "A" response,

$$P(A) = \frac{\exp(\phi a_A)}{\sum_k \exp(\phi a_k)}, \tag{2}$$

where a_k is the activation of category node k, and ϕ is a response scaling parameter.

ALCOVE adjusts its dimensional attention and association weights during learning via gradient descent, and thus learns to attend to diagnostic stimulus dimensions and ignore irrelevant dimensions. Attention weights are constrained to sum to unity across stimulus dimensions by a normalizing process. The rates at which associations and dimensional attention weights are learned are determined by learning rate parameters, λ_e and λ_α, respectively.

7.2. ATRIUM: A Mixture-of-Experts Model of Categorization

ATRIUM (presented schematically in Figure 10) instantiates a modular architecture that is broadly consistent with theories that assume knowledge

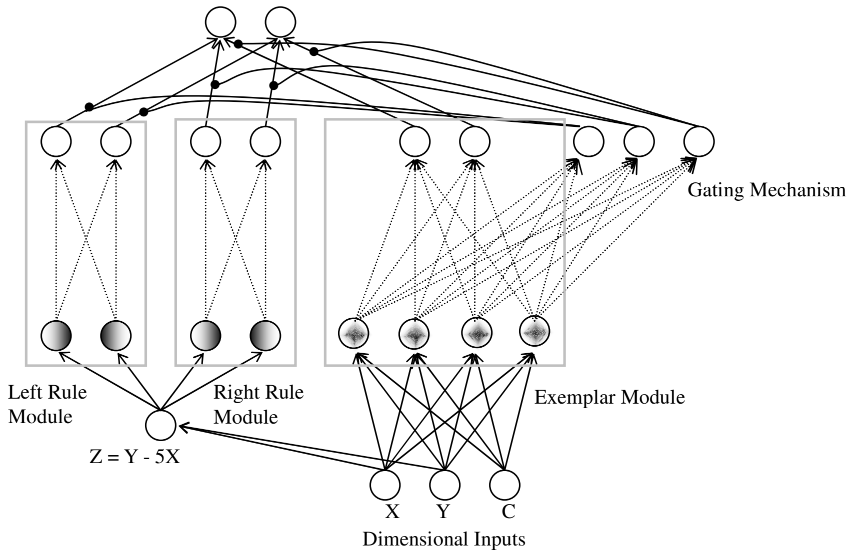


Figure 10: Schematic representation of the architecture of ATRIUM. The model consists of two rule modules instantiating the left and right partial boundaries and an exemplar module. Rule modules depicted in the figure are from Experiment 1. Dotted lines represent connections with learned weights. Figure adapted from Erickson, M. A., and Kruschke, J. K., Rules and exemplars in category learning, *Journal of Experimental Psychology: General*, *127*, 107–140, 1998, published by the American Psychological Association, used with permission.

to have a highly heterogeneous, fractionated structure (e.g., diSessa, 1988; diSessa, Gillespie, & Esterly, 2004; Lewandowsky, Little, & Kalish, 2007). Modules independently learn to categorize stimuli and compete to produce output. A gating mechanism (Jacobs, Jordan, Nowlan, & Hinton, 1991; Jacobs, 1999) adjudicates between the modules, and determines the contribution each makes to the final overall response.

The implementation of ATRIUM for Experiment 2 is similar to that of Erickson and Kruschke (1998). However, the same is not true for Experiment 1 because ATRIUM defines rules as orthogonal to a single psychological dimension. We therefore had to adapt the rule modules to accommodate the multidimensional partial boundaries from Experiment 1 (described by $y = 5x - 150$ and $y = 5x - 650$, respectively, in Figure 3). Rules were represented along a compound psychological dimension, z, defined by $z = y - 5x$ (cf. Yang & Lewandowsky, 2004). The compound dimension z describes a way of computationally representing the result of explicitly comparing the diameters of the inner and outer circles, which is commensurate with participants' written protocols whilst also respecting the results of the scaling study (see Appendix B).

Thus, the versions of ATRIUM used to fit the data from Experiments 1 and 2 incorporated both rule and exemplar representations; separate rule modules represented the left and right partial boundaries. For Experiment 1, both rule modules received input along the compound z dimension. For

Experiment 2, the left and right rule modules received input along the y and x dimensions, respectively. For both experiments, the exemplar module received input from the x, y, and context dimensions.

7.2.1. Rule Modules

Rule modules represent decision boundaries that bisect the category space into different response regions. Each rule module contains a pair of input nodes whose activation is determined by the positioning of stimulus input relative to the boundary represented by that module. The extent to which input i along dimension s activates one of the input nodes from pair j (denoted by $+$ and $-$ subscripts, respectively) is determined by a sigmoid function,

$$a_{r_{sj+}} = \frac{1}{1 + \exp\left[-\gamma\left(i_s + \beta_s\right)\right]}, \tag{3}$$

and the extent to which the other input node of pair j, $a_{r_{sj-}}$, is activated by input i is $1 - a_{r_{sj+}}$. In Equation 3, the value of γ is proportional to the standard deviation of the perceptual or criterial noise associated with the rule, thus determining the rule "sharpness". β_s determines the position of the rule boundary along dimension s. During training, response regions on either side of the rule boundary are associated with category output nodes by gradient descent on error. Category outputs from each module are fed to the gating mechanism as candidate category responses. The rates at which the rule modules learn these input-to-category associations are governed by learning rate parameters, λ_L and λ_R for the modules that instantiate the left and right partial boundaries, respectively.

7.2.2. Exemplar Module

The exemplar module is a standard implementation of ALCOVE except that instead of being used to directly compute responses, activation of the module's category nodes are fed forward to the gating mechanism as candidate responses.

7.2.3. Gating Mechanism

ATRIUM's gating mechanism determines the extent to which each module contributes to the model's final response, giving rise to what Erickson and Kruschke (1998) term *representational attention*, which refers to the capacity to associate specific stimuli with different representational modules.

The gating mechanism relies on the same exemplar representation as the exemplar module, and over the course of training, learns to associate exemplars to modules, such that the module that is best suited to categorizing a particular training exemplar will contribute most to the model's overall response. A learning rate parameter, λ_g, determines how quickly the exemplar-to-module associations are learned.

Stimulus activation of the gating mechanism is driven by exemplar similarity such that similar stimuli are categorized by a common module. For example, if a novel stimulus is highly similar to stored exemplars that were categorized using the left rule module, that module will also categorize the novel stimulus. Because the behavior of the gating mechanism is determined by exemplar similarity, it follows that representational attention is fundamentally dependent on the underlying distribution of dimensional attention. Given the known relationship between the distribution of dimensional attention and the similarity structure of the task (Nosofsky, 1987), changes in dimensional attention may indirectly elicit a change in the distribution of representational attention. We identify this as a possible mechanism of recoordination.

Formally, the computation of gains for each module is described by,

$$g_m = \frac{\exp\left(\phi_g a_{g_m}\right)}{\sum_m \exp\left(\phi_g a_{g_m}\right)}, \tag{4}$$

where a_{g_m} is the activation of the gate node for module m, and ϕ_g is a scaling parameter that defines how sensitive the gating mechanism is to differences in gate node activations.

Equation 4 normalizes the scaled gate activations to sum to unity, expressing the relative contribution of each module to the final model output in terms of module choice probabilities. The model's response is computed by weighting the candidate response generated by each module by its gain, and summing across modules. Thus, the probability that the model will generate an "A" response is given by,

$$P\left(A\right) = \sum_m g_m \frac{\exp\left(\phi a_{m_A}\right)}{\sum_k \exp\left(\phi a_{m_k}\right)}, \tag{5}$$

where m indexes the modules, and k indexes the category nodes in each module. Thus, a_{m_A} denotes the activation of the A category node in module m. The parameter ϕ is a response scaling parameter that determines how sensitive the model is to differences in category node activation in each module.

7.2.4. Modeling Knowledge Restructuring

In the Introduction, we discussed two knowledge restructuring processes: relearning and recoordination. Relearning is available to both ALCOVE and ATRIUM via incremental, error-driven adjustment of various weight matrices. Thus, the mechanism of relearning is identified with the error-driven learning mechanism common to both models. It is less obvious what mechanism(s) might drive recoordination. Recall that we define recoordination as a change in the way existing subsets of knowledge are applied to the task. Although our definition is most readily identified with a modular architecture, it is possible for a single-module model to implement a recoordination-like process. For example, within ALCOVE, recoordination could operate via changes in exemplar specificity, c in Equation 1. By broadening or focusing the exemplar generalization gradient, the quantity of exemplar-to-category associations recruited to categorize a given stimulus will either increase or decrease. Alternatively, recoordination could affect decision-level performance; changes in ϕ would influence whether the same stimulus would elicit a more (or less) deterministic response. Finally, recoordination might be realized in ALCOVE via a change in its dimensional attention weights. Changes in these weights will modify the pattern of exemplar activation in response to a given stimulus, which will in turn modify the pattern of exemplar-to-category association weights used to categorize the stimulus. In our modeling, all of these possible mechanisms (including possible interactions between them) were made available to ALCOVE by re-estimating these parameters anew for each transfer test. By the same token, recoordination could be implemented in a variety of ways in ATRIUM. However, we focus on the effects of changing dimensional attention weights as a means of influencing the behavior of ATRIUM's gating mechanism. With one exception, which required an additional change in gating "decisiveness", ϕ_g, we found that changes in dimensional attention in ATRIUM were sufficient to induce subsequent changes in representational attention, and thus knowledge restructuring.

7.3. Experiment 1 Modeling

To test the first modeling prediction based on a recoordination account of our data, we separately fit ALCOVE and ATRIUM, to the Initial-KP and

Initial-CI group transfer data from Experiment 1. ALCOVE had four parameters, all of which were also incorporated into ATRIUM: The exemplar specificity, c, the response scaling parameter, ϕ, and two learning rates, λ_e and λ_α. In addition to these parameters, ATRIUM had the scaling parameter for the gating mechanism, ϕ_g, the noise parameter for the rule modules, γ, and learning rates for the rule modules, λ_L and λ_R, and the gating mechanism, λ_g.

For all model fits, stimuli were coded by linearly transforming their values into the range [0,1]. Context was coded as a binary dimension represented by a 0 or a 1. Parameters were estimated using maximum likelihood methods assuming a binomial model. The standard simplex algorithm (Nelder & Mead, 1965) was used to minimize the negative log likelihood,

$$-\ln L = -\left[\sum_i f_i \ln(p_i) + (n - f_i)\ln(1 - p_i)\right], \tag{6}$$

where p_i is the predicted probability with which item i was assigned to category A, f_i is the observed frequency of the ith item being assigned to category A, and n is the number of observations. This statistic was used to compute Akaike Information Criterion (AIC; Akaike, 1974) values for model comparison purposes. The AIC is defined as,

$$AIC = -2\ln L + 2k, \tag{7}$$

where k is the number of free parameters. The latter term penalizes model complexity, thus permitting comparisons between models with different numbers of free parameters.

For Experiment 1, the effect of the hint was modeled by adding to each model, a free parameter, α_c, which determined the attentional loading onto the context dimension following presentation of the hint in each phase. Specifically, the distribution of dimensional attention after presentation of the hint was determined by estimating the attentional loading onto context, and renormalizing the remaining weights according to their learned values according to:

$$\alpha'_s = (1 - \alpha_c)\frac{\alpha_s}{\sum_{k \neq c} \alpha_k}, \tag{8}$$

where α_s denotes the attention weight on dimension s, prior to administration of the hint and α'_s is the corresponding weight after the hint.

Thus, ATRIUM was fit to the data from each transfer phase in two steps: In the first step, models were fit directly to the first test of that transfer phase by estimating all parameters except for α_c, which was learned as per the other attention weights. For the second test within each phase, only α_c was estimated while all other parameters were held constant. By contrast, we re-estimated all of ALCOVE's parameters anew for each transfer test. This allowed us to examine the extent to which recoordination might be modeled within a single-module architecture whilst also giving ALCOVE the best possible chance of fitting the data. Incidentally, ALCOVE failed quite badly if the attentional loading onto context, α_c, was the only parameter re-estimated for Tests 2 and 4.

Preliminary exploration of the parameter space revealed that in ATRIUM, ϕ_g and λ_g could be yoked across all simulations without loss of fit. Similarly, within the Initial-KP group, c, λ_e, and λ_R could be yoked across the phase-wise fits; within the Initial-CI group, the c and ϕ parameters could be yoked.

Best-fitting parameters for ALCOVE and ATRIUM are reported in Table 1. Fit statistics and attention weights for each test are reported in Table 2. Although it was not minimized when conducting fits, we present a trimmed root mean squared deviation (RMSD) as a descriptive fit statistic. To compute the trimmed-RMSD, the 10 worst fitting data points from each test's set of 160 were excluded.

7.3.1. Prediction 1: Model Fits

On the basis of AICs, it is clear that ATRIUM provides a superior overall account of the data from Experiment 1. It is interesting to note that ALCOVE provides a superior fit for the Initial-CI group on Test 4. This is entirely due to differences in the learned weights within the two models. Although one might expect ATRIUM to provide a better fit on the grounds that it contains ALCOVE as a special case, this does not follow, as the models learned the task in quite different ways. Since no new additional learning occurred between Tests 3 and 4, it follows that the fit to Test 4 reflects the capabilities of the models given knowledge they had acquired during the previous learning phases. We also note that ALCOVE only provides a better fit to the Test 4 data if the c and ϕ parameters are re-estimated along with the full set of attention weights. Re-estimating only the attention weights produced a poorer fit ($AIC = 2031$).

We present ALCOVE's best fitting predictions to performance on the first transfer test for both the Initial-CI and Initial-KP groups in Figure 11. It

Table 1: Best fitting parameters for both models for the data from all transfer tests in Experiment 1 for the Initial-CI and Initial-KP groups. Yoked ATRIUM parameters straddle columns.

<table>
<tr><th colspan="9">ALCOVE</th></tr>
<tr><th></th><th colspan="4">Initial-CI Group</th><th colspan="4">Initial-KP Group</th></tr>
<tr><th></th><th>Test 1</th><th>Test 2</th><th>Test 3</th><th>Test 4</th><th>Test 1</th><th>Test 2</th><th>Test 3</th><th>Test 4</th></tr>
<tr><td>c</td><td>6.31</td><td>5.39</td><td>2.47</td><td>1.24</td><td>7.70</td><td>3.24</td><td>3.54</td><td>2.01</td></tr>
<tr><td>ϕ</td><td>9.69</td><td>8.92</td><td>2.77</td><td>12.55</td><td>1.29</td><td>2.82</td><td>.94</td><td>.62</td></tr>
<tr><td>λ_e</td><td>.09</td><td>-</td><td>.01</td><td>-</td><td>.19</td><td>-</td><td>.15</td><td>-</td></tr>
<tr><td>λ_α</td><td>.01</td><td>-</td><td>.02</td><td>-</td><td>.0001</td><td>-</td><td>.0001</td><td>-</td></tr>
<tr><td>α_x</td><td>-</td><td>.50</td><td>-</td><td>.65</td><td>-</td><td>.49</td><td>-</td><td>.62</td></tr>
<tr><td>α_y</td><td>-</td><td>.50</td><td>-</td><td>.35</td><td>-</td><td>.38</td><td>-</td><td>.38</td></tr>
<tr><td>α_c</td><td>-</td><td>.00</td><td>-</td><td>.00</td><td>-</td><td>.13</td><td>-</td><td>.00</td></tr>
<tr><th colspan="9">ATRIUM</th></tr>
<tr><td>c</td><td colspan="4">1.14</td><td colspan="4">0.56</td></tr>
<tr><td>ϕ</td><td colspan="4">7.35</td><td colspan="2">5.74</td><td colspan="2">3.36</td></tr>
<tr><td>λ_e</td><td colspan="2">.90</td><td colspan="2">.98</td><td colspan="4">1.03</td></tr>
<tr><td>λ_α</td><td colspan="2">3.24</td><td colspan="2">3.45</td><td colspan="2">1.30</td><td colspan="2">2.12</td></tr>
<tr><td>ϕ_g</td><td colspan="8">2.52</td></tr>
<tr><td>λ_L</td><td colspan="2">2.60</td><td colspan="2">2.71</td><td colspan="2">2.79</td><td colspan="2">.01</td></tr>
<tr><td>λ_R</td><td colspan="2">2.59</td><td colspan="2">3.01</td><td colspan="4">2.03</td></tr>
<tr><td>λ_g</td><td colspan="8">.06</td></tr>
<tr><td>γ</td><td colspan="2">1.42</td><td colspan="2">.92</td><td colspan="2">1.74</td><td colspan="2">1.08</td></tr>
<tr><td>α_c</td><td>-</td><td>.24</td><td>-</td><td>.11</td><td>-</td><td>.14</td><td>-</td><td>.27</td></tr>
</table>

Table 2: Experiment 1 fit statistics and attention weights for ALCOVE and ATRIUM for each transfer test for the Initial-CI and Initial-KP groups. Lowest AIC values for each test are presented in bold.

	Test	AIC	$-lnL$	$RMSD$	α_x	α_y	α_c
ALCOVE							
Initial-CI	1	2035.20	1013.59	.1351	.49	.47	.04
	2	2304.46	1148.23	.1521	.50	.50	.00
	3	2417.52	1204.76	.1809	.48	.47	.06
	4	**1899.24**	945.62	.1102	.65	.35	.00
Initial-KP	1	1352.80	672.42	.2169	.36	.40	.24
	2	1320.98	656.49	.1518	.49	.38	.13
	3	1025.26	508.63	.1159	.45	.43	.12
	4	1363.56	677.78	.1805	.62	.38	.00
ATRIUM							
Initial-CI	1	**1820.10**	901.05	.0963	.18	.69	.14
	2	**1879.28**	938.64	.0890	.17	.59	.24
	3	**1653.76**	817.88	.0902	.19	.56	.24
	4	1943.70	970.85	.1188	.24	.65	.11
Initial-KP	1	**967.76**	474.88	.1144	.29	.37	.34
	2	**1235.20**	616.60	.1159	.40	.46	.14
	3	**982.94**	482.47	.1064	.37	.51	.13
	4	**1134.38**	566.19	.1050	.31	.42	.27

is clear that ALCOVE fails to capture many of the qualitative aspects of performance. ALCOVE overpredicts the extent of P(A) responses in Areas 3 and 4 for the Initial-CI group, and is unable to reproduce the effect context had on responding of the Initial-KP group. We conclude that ATRIUM provides the best account of the data from Experiment 1, and do not discuss ALCOVE any further. The superior fit of ATRIUM to the data of Experiment 1 is thus consistent with the first prediction made by a recoordination account of the data.

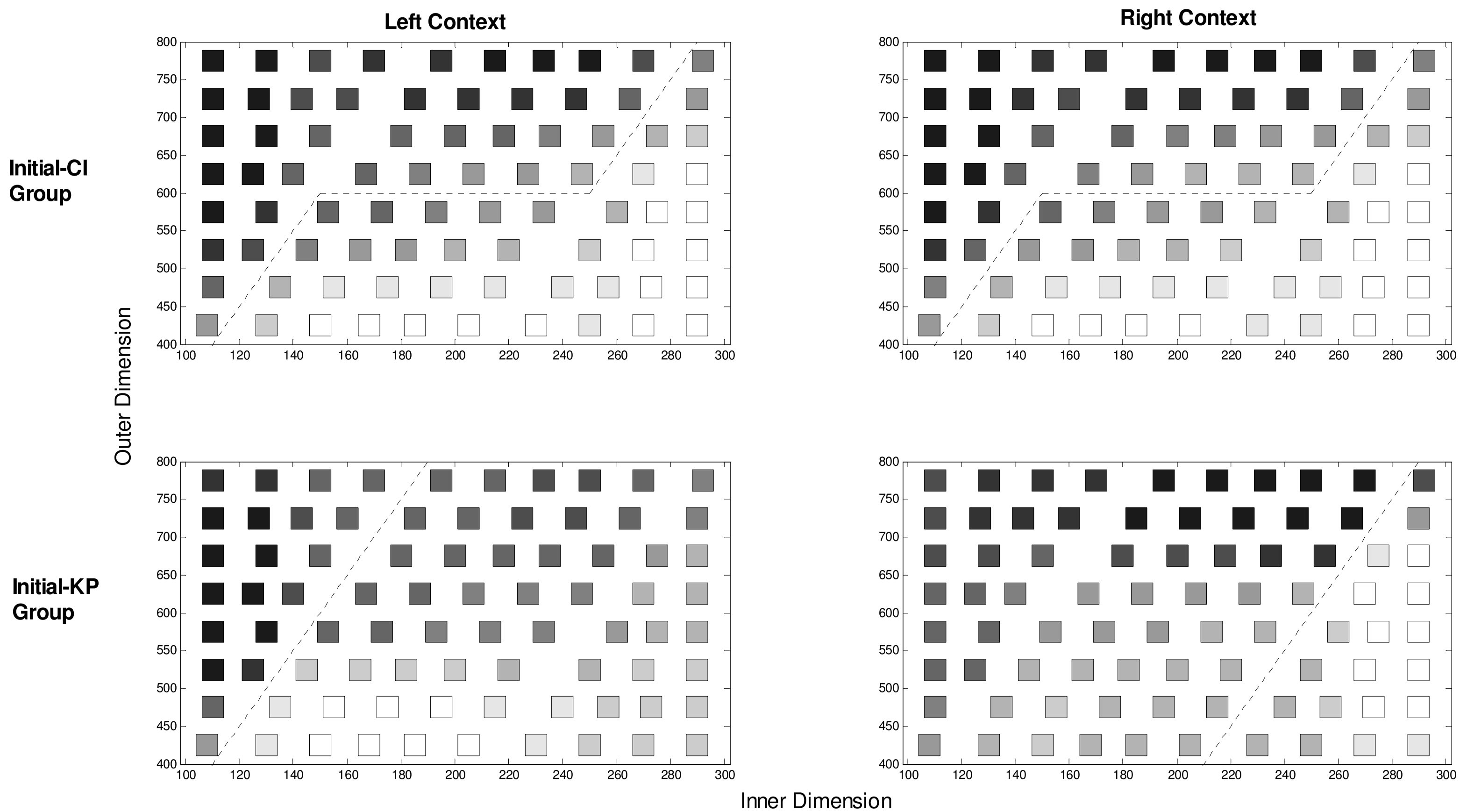


Figure 11: ALCOVE fit to the results from the first transfer test 1 of Experiment 1 for the Initial-CI and Initial-KP groups (top and bottom panels, respectively). Strategy-appropriate category boundaries are included for emphasis. Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

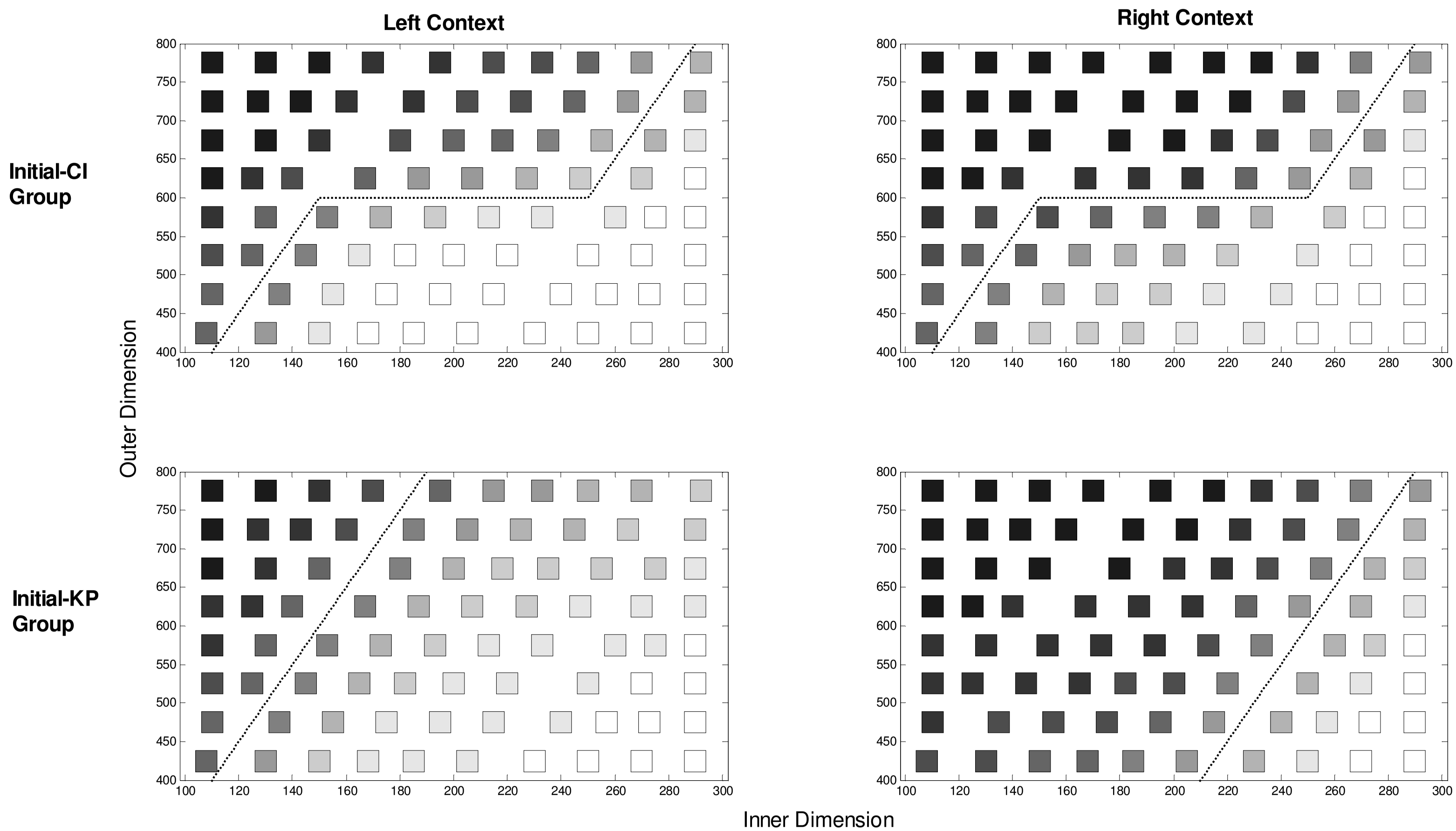


Figure 12: ATRIUM fit to the results from the first transfer test 1 of Experiment 1 for the Initial-CI and Initial-KP groups (top and bottom panels, respectively). Strategy-appropriate category boundaries are included for emphasis. Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

As can be seen in Figure 12, ATRIUM provides a satisfactory account of the performance of both groups in the initial test phase. The model's capacity to track people's performance across the entire experiment is apparent from Figure 5, which plots predicted context sensitivity alongside the data. Regardless of whether the model initially applied the KP or CI strategy, modulation of the attentional loading on the context dimension was sufficient to shift response strategy. Moreover, the notion that ATRIUM, like the participants in Experiment 1, was able to recover its initial categorization strategy, even after having previously restructured its knowledge, is supported by high item-wise correlations (computed as per Experiment 1) between performance in transfer Tests 1 and 4, regardless of whether the model was fitting the data from the Initial-KP group ($r = .91$) or the Initial-CI group ($r = .89$).

It is clear that ATRIUM provides a reasonable account of the behavioral data from Experiment 1. The substantial changes in the model's behavior in response to adjustment in the attentional loading onto the context dimension (i.e., the hint) imply a tight link between dimensional and representational attention. Table 2 illustrates the close coupling between attention to context and the manifest categorization strategy used by the model (cf. Yang & Lewandowsky, 2004). When applying the CI strategy (Initial-CI Group, Tests 1 and 4; Initial-KP Group, Tests 2 and 3), the attentional loading on context was relatively low. In contrast, when applying the KP strategy (Initial-CI Group, Tests 2 and 3; Initial-KP Group, Tests 1 and 4), attention to context was higher.

Although suggestive, the fact that ATRIUM can track behavioral changes across four transfer tests does not speak to whether recoordination drove knowledge restructuring in the model. To address the remaining three modeling predictions, we must investigate how changes in dimensional attention relate to response-level changes. The first step is to assess how different distributions of dimensional attention affect behavior at the level of the gating mechanism.

7.3.2. Prediction 2: Analysis of Gain Profiles

The behavior of ATRIUM's gating mechanism is characterized by the module gains computed for each transfer stimulus. For a given stimulus, module gains are a vector describing the probability that each module will categorize the stimulus. If changes in dimensional attention elicit subsequent changes in representational attention (i.e., if knowledge restructuring is based on recoordination), it follows that the patterns of gains should be strategy-dependent.

To explore the link between dimensional attention and overt strategic behavior, module gains for transfer stimuli presented in each context were aggregated for each diagnostic area. If dimensional attention shifts resulted in differences in how ATRIUM applied its partial knowledge to the task (i.e., knowledge restructuring), each strategy should be associated with a distinct gains profile. Table 3 presents gains averaged across tests to which a common strategy was applied (i.e., Tests 1 and 4, and Tests 2 and 3) for the Initial-KP and Initial-CI groups. Given ALCOVE's inability to model task performance, it is perhaps not surprising that ATRIUM did not rely on the exemplar module to perform the task (cf. Yang & Lewandowsky, 2004).

Across groups, strategy-specific regularities in gain profiles are apparent. When applying the CI strategy, the dominant module in Areas 2 and 3 (the diagnostic region of the stimulus space) is consistent across contexts. ATRIUM relies on the right rule module in Area 2, whereas the left rule dominates in Area 3. Conversely, when the model is applying the KP strategy, the dominant module depends on the context in which transfer items are presented. For stimuli presented in the left context, the left rule module dominates, for stimuli presented in the right context, the right rule module drives responding. The strategy-specific gain profiles thus confirm that changes in representational attention can arise from dimensional attention shifts, implying a causal role of selective dimensional attention in recoordination of knowledge.

7.3.3. Predictions 3 & 4: Analysis of Partial Knowledge

We have established that ATRIUM invokes a recoordination process to model the knowledge restructuring in Experiment 1. Table 3 shows that changes in dimensional attention result in changes in the way the model gates its modular rule-based knowledge. In both cases, ATRIUM performs the task in a rule-driven way. Accordingly, we analyze the patterns of learned weights in ATRIUM's rule modules to examine whether different strategies are reliant on common partial knowledge (Prediction 3), and also whether partial knowledge used within a given group is sensitive to changes in overt response strategy (Prediction 4).

Recoordination predicts that strategy changes arise from changes in the coordination of partial knowledge, not the content of that knowledge. The pattern of gains in Table 3 suggests that people are shifting between distinct strategy-dependent ways of coordinating their rule-based knowledge.

Table 3: Gains for transfer stimuli in presented in each context, aggregated across items within each diagnostic area, for both Initial-CI and Initial-KP groups from Experiment 1. For each group, gains are averaged across transfer tests on which a common strategy was applied.

		Initial-CI Group					
		Left Context			Right Context		
Tests	Area	Left Rule	Right Rule	Exemplar	Left Rule	Right Rule	Exemplar
1 and 4	1	.66	.29	.06	.37	.56	.07
	2	.31	.61	.08	.07	.89	.04
	3	.90	.06	.04	.63	.29	.08
	4	.58	.34	.08	.30	.64	.06
2 and 3	1	.78	.17	.05	.22	.70	.07
	2	.53	.39	.09	.03	.94	.03
	3	.94	.03	.03	.39	.52	.09
	4	.71	.21	.07	.16	.78	.06
		Initial-KP Group					
1 and 4	1	.89	.10	.01	.24	.75	.01
	2	.66	.33	.01	.04	.95	.00
	3	.91	.08	.01	.28	.71	.01
	4	.68	.30	.01	.08	.92	.01
2 and 3	1	.81	.18	.01	.52	.47	.01
	2	.36	.63	.01	.10	.89	.01
	3	.85	.14	.01	.56	.43	.01
	4	.40	.58	.01	.14	.85	.01

If this is the case, it follows that the content of those rules should be very similar even between groups using different strategies. We addressed this question by compiling, for each group in each phase of the experiment, the learned rule module weights into separate vectors. We averaged the weights across simulation runs for each group, then computed the cosine of the angle formed between the group averaged vectors. This cosine similarity measure is bounded in the range [-1, 1], is closely related to the Pearson correlation, and is interpreted in much the same way. The cosines reflect similarity in the content of the rule modules, and thus indicate the extent to which the Initial-KP and Initial-CI groups relied on common partial knowledge despite implementing quite different response strategies. There were near perfect between-group correlations between learned weights at the end of Phase 1 for both the left and right rule modules (both $\cos\theta$s $> .99$). Correlations were similar after learning in Phase 2 for both rule modules (both $\cos\theta$s $> .99$). Thus, it is clear that the Initial-CI and Initial-KP groups were reliant on common partial knowledge despite the large differences in overt responding. We then confirmed that our results were not an artifact of averaging weights before computing cosine similarity. Computing cosine similarity for each simulation run, then averaging the cosines yielded virtually identical results (for both modules in both phases, $\cos\theta$s $> .94$). We further confirmed that these correlational analyses were sensitive to in-principle differences in rule content: An analogous analysis on the rule weight matrix from the Complete Reversal simulation reported in the Introduction yielded a strong negative correlation between the Initially learned rule weights and those after relearning of the reversal shift, $\cos\theta = -.90$, reflecting the clear weight changes in the rule module for that simulation (see Appendix A for details).

The between-group weight correlations do not address how partial rule-based knowledge within a group is affected by knowledge restructuring. We therefore performed similar correlational analyses on the learned weight matrices of the rule modules across phases within each group. If rule content is invariant across strategies used within a group, it implies that the strategies are reliant on common elements of partial knowledge. For each group, we computed the between-phase correlation between weights in the left and right rule modules. For the Initial-KP group, the content of each rule module was largely unaffected by knowledge restructuring; correlations in both the left and right rule modules were very high, (both $\cos\theta$s $> .99$). The same pattern was observed for the Initial-CI. The content of the left and right rule modules was qualitatively unchanged despite the intervening knowledge restructuring (both $\cos\theta$s $> .99$). Again, we confirmed that similar results obtained when the cosines from each simulation run were averaged (for both modules in both groups $\cos\theta$s $> .92$). Thus, the fourth prediction based on recoordination was also confirmed: The content of partial knowledge is unaffected by overt strategy use. The same rule-based knowledge underpinned the CI and KP strategies in Experiment 1.

7.4. Experiment 2 Modeling

For these fits, the parameters for ALCOVE were the same as in Experiment 1 whereas for ATRIUM, a single rule learning rate parameter, λ_{LR}, turned out to be sufficient. Because hints were administered before commencement of the task in Experiment 2, we freely estimated the initial vector of dimensional attention (i.e., α_x, α_y, and α_c). Because dimensional attention tended to load heavily onto a single dimension, we were unable to model knowledge restructuring using the re-normalization method described by Equation 8. We therefore re-estimated the full set of attention weights anew for each transfer test. Note that we re-estimated all of ALCOVE's parameters for each transfer test, like in Experiment 1. For ATRIUM, only the dimensional attention parameters were re-estimated for Tests 2 and 4 with one exception; for the CI-first condition in Test 4, we also had to re-estimate the gate decisiveness parameter, ϕ_g. Although this was, strictly speaking, not an attentional parameter, the modeling results still accord with a recoordination account: ϕ_g only controls the selection of candidate response modules, it does not affect their content.

One final comment about the modeling concerns performance of the CI-first condition in the right context of Test 2. In the CI-first condition, there were individual differences in performance in the right context of Test 2 relating to where along the x dimension the vertically oriented right partial boundary was placed. One rule, used by 45% of CI-first participants, was identical to the vertically oriented right partial boundary depicted in Figure 6. The other rule, used by 55% of CI-first participants, was also vertically oriented, but positioned at the *midpoint* along the x dimension. This unexpected rule, which was restricted to the right context of this single transfer test, arose from an ambiguity in the wording of the hint. The hint describing the KP strategy stated that category membership in the right context could be determined by exclusively considering the x dimension. For the CI-first participants though, the x dimension served an additional purpose beyond determining category membership; namely, determining whether the x or y

dimension was ultimately applicable for a given stimulus. Stimuli to the left of the vertical midline of the category space could be categorized using the y dimension, those to the right of the vertical midline could be categorized using the x dimension. Many of the CI-first participant used this "midline" rule in preference to the intended vertical rule. This did not impose any difficulties in interpreting strategy use: Performance of this group in Test 2 was still governed by a variant of the KP strategy; a rectangle height rule was used in the left context, whereas a bar position rule was used in the right context. To model this variation in behavior, we generated predictions from two parallel versions of ATRIUM to the Test 2 data; the versions differed only in terms of the placement of the rule along the x dimension (i.e., differences were restricted to the β_s parameter in Equation 3). All other parameters assumed the same values across different model versions. Further, the same learned weights characterized the vertical rule, regardless of where it was placed. Thus, we do not consider the "midline" rule as distinct from the intended vertical rule — from a modeling perspective, it is the *same* rule, shifted along the x dimension. The predictions of each version of the model were then weighted according to the prevalence of the two types of vertical rules in the data (i.e., $P(A) = .55_{\text{Midline}} + .45_{\text{Intended}}$, where Midline and Intended describe the predictions of models instantiating the two positions of the rule along the x dimension). Test 2 parameters for the CI-first condition were optimized with respect to the weighted predictions.

7.4.1. Prediction 1: Model Fits

Table 4 displays best fitting parameters for ALCOVE and ATRIUM to the data from Experiment 2. Fit statistics for the CI-first and KP-first conditions in each transfer test are shown in Table 5.

The results are clear-cut. In each transfer test, ATRIUM provides a substantially better quantitative fit to the data, despite the fact that ALCOVE's c and ϕ parameters were re-estimated anew for each test. It is also noteworthy that ATRIUM provides an excellent fit to the CI-first data from Test 2, where people placed the vertical rule in one of two places along the x dimension. The good fit is noteworthy because the two versions of the model used to fit those data were reliant on a common set of learned weights (i.e., weights learned after the initial training phase). This implies that the content of the vertical rule was the same, regardless of where along the x dimension participants placed it, justifying our earlier claim that the "midline" rule should not be considered as different from the intended rule.

Table 4: Best fitting parameters for both models for the data from all transfer tests in Experiment 2 for the CI-first and KP-first conditions.

	ALCOVE							
	CI-first Condition				KP-first Condition			
	Test 1	Test 2	Test 3	Test 4	Test 1	Test 2	Test 3	Test 4
c	10.00	3.66	3.74	12.32	7.69	15.00	10.00	8.17
ϕ	1.14	1.30	1.85	1.04	1.86	.73	1.11	1.31
λ_e	.39	-	.10	-	.23	-	.46	-
λ_α	.00	-	.00	-	.0001	-	.001	-
α_x	.80	.44	.39	.82	.38	.77	.83	.38
α_y	.17	.18	.15	.18	.20	.23	.16	.14
α_c	.03	.38	.46	.01	.42	.00	.00	.49
	ATRIUM							
c	12.41	-	14.12	-	3.70	-	15.00	-
ϕ	10.18	-	2.29	-	3.85	-	1.62	-
λ_e	.01	-	.01	-	.004	-	.01	-
λ_α	.00	-	.00	-	.001	-	.001	-
ϕ_g	2.06	-	.07	.35	7.97	-	11.83	-
λ_{LR}	1.74	-	.00	-	4.79	-	1.65	-
λ_g	.05	-	2.75	-	.002	-	.004	-
γ	4.31	-	9.99	-	3.08	-	4.69	-
α_x	.97	.20	.00	.97	.19	1.00	.97	.22
α_y	.00	.11	.00	.03	.07	.00	.00	.16
α_c	.04	.69	1.00	.01	.74	.00	.03	.62

Table 5: Experiment 2 fit statistics for ALCOVE and ATRIUM for each transfer test for the CI-first and KP-first conditions. Lowest AIC values for each test are presented in bold. Final learned attention weights are not presented in this Table because they did not deviate much from the initial values set by the hint.

	Test	AIC	$-lnL$	$RMSD$
		ALCOVE		
CI-first	1	2469.68	1228.84	.0973
	2	2574.68	1283.34	.1261
	3	2391.46	1189.73	.0687
	4	2318.84	1155.42	.1014
KP-first	1	1859.86	923.93	.0887
	2	3164.90	1578.45	.1762
	3	2631.20	1309.60	.1041
	4	2430.96	1211.48	.0894
		ATRIUM		
CI-first	1	**2182.52**	1081.26	.0447
	2	**2227.36**	1111.68	.0612
	3	**2312.50**	1146.25	.0496
	4	**1966.76**	980.38	.0490
KP-first	1	**1251.32**	615.66	.0265
	2	**2233.76**	1114.88	.0590
	3	**2100.44**	1040.22	.0360
	4	**2138.94**	1067.47	.0416

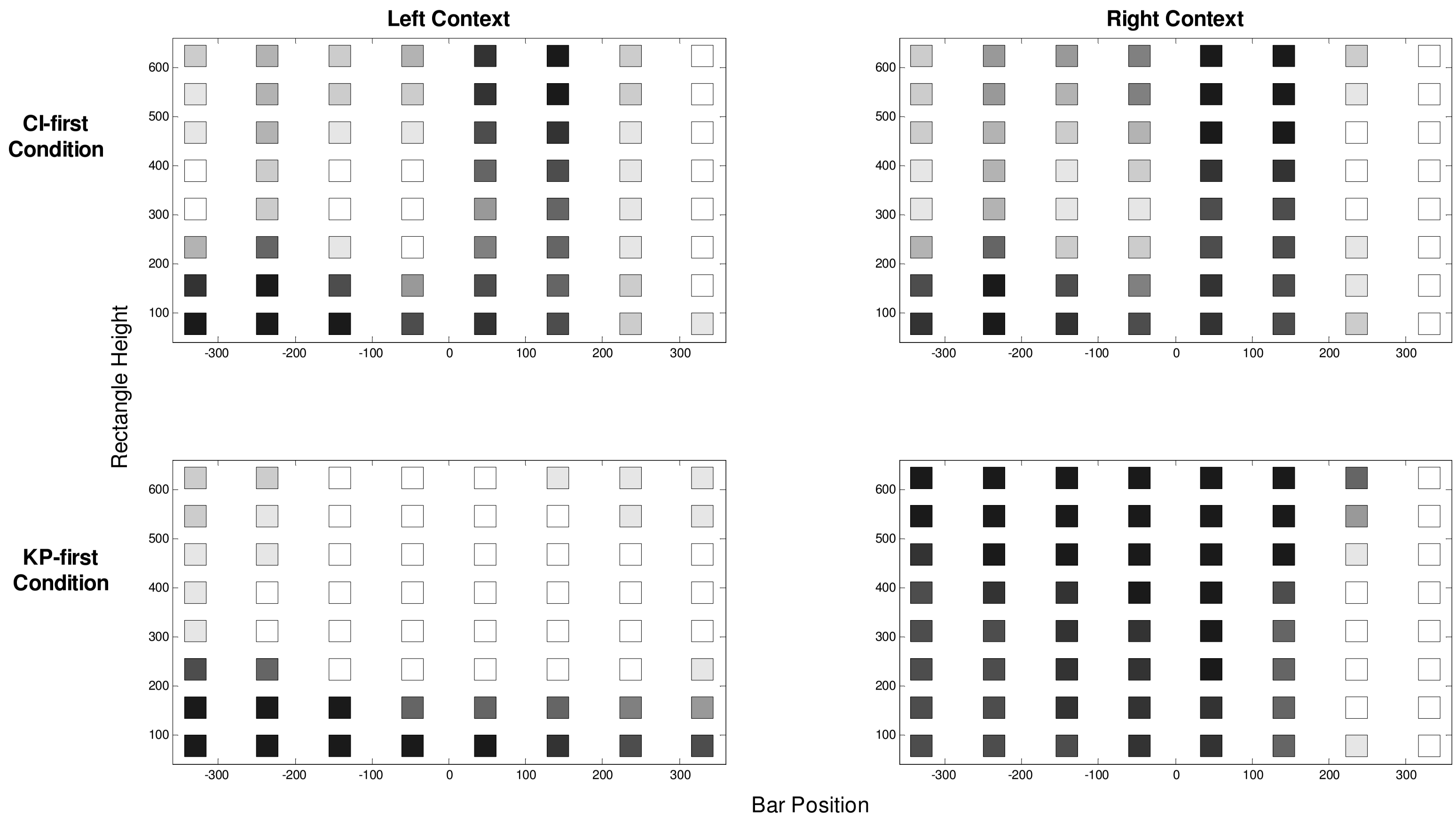


Figure 13: ALCOVE fit to the results from the first transfer test 1 of Experiment 2 for the CI-first and KP-first conditions (top and bottom panels, respectively). Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

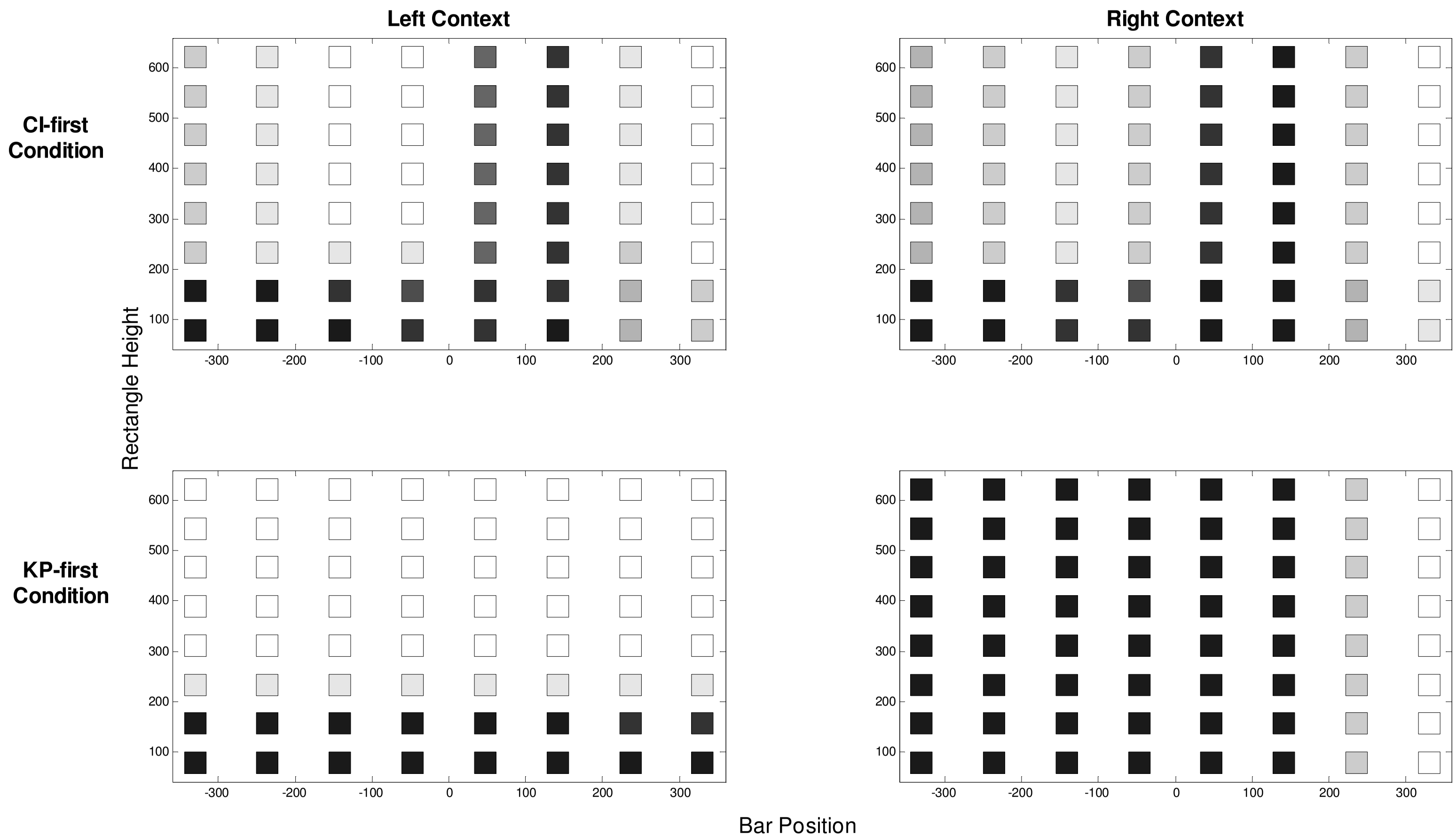


Figure 14: ATRIUM fit to the results from the first transfer test 1 of Experiment 2 for the CI-first and KP-first conditions (top and bottom panels, respectively). Darker levels of shading correspond to higher P(A). Shading varies in steps of .1.

ALCOVE and ATRIUM's Test 1 predictions for each group are shown in Figures 13 and 14, respectively. It is clear that ALCOVE misses key qualitative features of the data. For one, ALCOVE's CI-first condition predictions for mid-range values along the y dimension are inconsistent with the data; namely, the model predicts higher P(A) than are present. For ALCOVE's fit to the KP-first condition, the model is unable to extrapolate in a rule-based fashion, and thus fails to fit the data (cf. Denton et al., 2008; Erickson & Kruschke, 2002b). By contrast, ATRIUM very closely matches the data in both CI-first and KP-first conditions. ATRIUM's ability to track performance across the entire experiment is illustrated in Figure 9. ATRIUM's predicted context sensitivity matches participant data almost exactly in both conditions. Finally, ATRIUM's recovery of the original strategy is further supported by high item-wise correlations for model predictions to Tests 1 and 4 for the CI-first ($r = .97$) and KP-first ($r = .98$) conditions.

Overall, it is clear that the model fitting results are consistent with the first prediction based on recoordination: A multiple-module model better accounts for the data of Experiment 2. Accordingly, we do not discuss ALCOVE further. We now examine the gain profiles of the CI-first and KP-first conditions to establish whether knowledge restructuring in ATRIUM was driven by a recoordination process.

7.4.2. Prediction 2: Analysis of Gain Profiles

Once again, module gains for transfer stimuli presented in each context were aggregated for each diagnostic quadrant (refer to Figure 6). If the changes in the distribution of dimensional attention resulted in recoordination, the CI and KP strategies should be characterized by distinct gain profiles. Table 6 presents gains averaged across tests to which a common strategy was applied (i.e., Tests 1 and 4, and Tests 2 and 3) for both conditions. As in Experiment 1, ATRIUM's behavior was dominated by its rule modules; the model only sparingly relied on its exemplar module.

As in Experiment 1, there are strategy-specific regularities in gain profiles. When applying the CI strategy, the dominant module in each quadrant is consistent across contexts. Specifically, for quadrants on the left-hand side of the space (Quadrants 1 and 4), the left (horizontal) rule dominates, whereas the right (vertical) rule dominates on the right-hand side of the space (Quadrants 2 and 3). For the KP strategy though, the dominant module is entirely context-dependent. The left (horizontal) rule is used for stimuli presented in the left context, and the right (vertical) rule is used for stimuli presented in

Table 6: Gains for transfer stimuli in presented in each context, aggregated across items within each quadrant, for both CI-first and KP-first conditions from Experiment 2. For each group, gains are averaged across transfer tests on which a common strategy was applied.

		CI-first Condition					
		Left Context			Right Context		
Tests	Quadrant	Left Rule	Right Rule	Exemplar	Left Rule	Right Rule	Exemplar
1 and 4	1	.84	.06	.10	.75	.11	.14
	2	.11	.77	.11	.06	.86	.08
	3	.12	.76	.12	.07	.84	.09
	4	.86	.05	.09	.77	.10	.13
2 and 3	1	.82	.06	.12	.07	.82	.11
	2	.75	.09	.16	.04	.89	.07
	3	.79	.08	.14	.06	.86	.08
	4	.85	.05	.10	.10	.77	.13
		KP-first Condition					
1 and 4	1	.93	.02	.05	.03	.91	.05
	2	.80	.07	.13	.00	.99	.01
	3	.89	.03	.08	.02	.94	.04
	4	.97	.00	.02	.02	.94	.04
2 and 3	1	.87	.06	.07	.80	.09	.11
	2	.11	.81	.09	.07	.88	.05
	3	.11	.80	.09	.07	.88	.05
	4	.87	.06	.07	.80	.09	.11

the right context. As in Experiment 1, ATRIUM's behavior is dominated by rule use, and the changes in gain profiles imply that the shifts in dimensional attention we used to simulate knowledge restructuring had their effects at the level of ATRIUM's gating mechanism.

7.4.3. Predictions 3 & 4: Analysis of Partial Knowledge

We again correlated rule module weights after each learning phase between the CI-first and KP-first conditions. After Phase 1 learning, weights in the rule modules for each condition were highly correlated (both $\cos\theta$s > .99). The same pattern held after Phase 2 learning (both $\cos\theta$s > .99). Averaging cosines computed for each simulation run produced nearly identical results (for both modules in both phases, $\cos\theta$s > .97). As in Experiment 1, it is clear that the CI-first and KP-first conditions were reliant on common partial knowledge despite the large differences in overt responding.

To analyze within-group performance, we correlated weights in the two rule modules within each condition, but between phases. For the CI-first condition, the content of each rule module was largely unaffected by knowledge restructuring; correlations in both the left and right rule modules were very high across phases, (both $\cos\theta$s > .99). The same pattern was observed for the KP-first condition (both $\cos\theta$s > .99). Averaging cosines computed for each simulation run produced nearly identical results (for both modules in both groups, $\cos\theta$s > .99). Thus, the content of partial knowledge in ATRIUM was unaffected by overt strategy use. As in Experiment 1, the same rule-based knowledge underpinned both strategies in Experiment 2.

8. General Discussion

Much research in learning and development is motivated by the fact that the structure of existing conceptual knowledge constrains acquisition of further knowledge. The starting point for the current paper was the observation that different model architectures permit different forms of knowledge restructuring, including one that has remained unexamined in the literature, namely recoordination. We investigated several novel predictions about recoordination experimentally using a knowledge-partitioning paradigm (Yang & Lewandowsky, 2004). Experiments 1 and 2 verified that knowledge restructuring can occur in the absence of additional training, and that strategy shifts are not necessarily constrained by choice of initial response strategy.

This fluid, symmetrical knowledge restructuring has not been reported previously, and is a clear empirical indicator of a recoordination process. We then confirmed that even after initial knowledge restructuring and subsequent supervised training on an alternative strategy, people can recover their original strategy in response to a further hint. These key findings are difficult to reconcile with a relearning account of knowledge restructuring, and they speak against the generality of single-module models of category learning. Some degree of functional heterogeneity (or modularity) must be incorporated into a theory of conceptual structure if it is to describe the representational dynamics of knowledge restructuring.

We further examined the nature of the recoordination process via computational modeling. Consistent with recoordination, the data from Experiments 1 and 2 were better fit by the multiple-module ATRIUM model. ATRIUM implemented both the knowledge partitioning and context insensitive strategies via rule use. Differences in the way ATRIUM's rule modules were coordinated determined the manifest strategy: When rule selection was driven by context, knowledge partitioning emerged; when selection was driven by an item's location in x-y space, the context insensitive strategy was used. Importantly, knowledge restructuring in ATRIUM was driven by a shift in dimensional attention, which in turn affected the way in which ATRIUM coordinated its rule-based knowledge. We further confirmed that rule content was a) consistent across groups implementing different strategies, and b) insensitive to changes in overt response strategy within a given group. Hence, large differences in response-level behavior need not imply qualitative changes in the content of task knowledge. Taken together, the modeling results imply that knowledge restructuring can profitably be described in terms of changes in the coordination of partial knowledge elements, and that the key to flexible retention of strategic alternatives may lie in the modular decomposition of task knowledge. Overall, our results are broadly consistent with theories of strategy use and development that assume that people have ongoing access to a number of different strategies at a given point in time (e.g., Dixon & Kelley, 2007; Rieskamp & Otto, 2006; Siegler, 1996, 1999; though see also, Stephen, Dixon, & Isenhower, 2009). However, we point out that while these theories tacitly assume that strategies are stored as integrated representational structures (e.g., the overall success of a strategy determines whether it is retained or discarded), recoordination implies that people do not retain strategies *per se*; instead, people retain the partial knowledge elements that can be used in an ad hoc manner to construct strategies. The extent to which theories of strategy selection must be modified to accommodate this finer grain of selective retention of knowledge remains to be seen.

8.1. Exemplar Representation & the Mechanism(s) of Recoordination

We modeled hint-induced knowledge restructuring as the interaction between two forms of attention: Attention to stimulus dimensions, and attention to different partial solutions to a task (i.e., representational attention)[7]. Specifically, we focused on whether shifts of dimensional attention could drive higher-level shifts in representational attention, and thus serve as a plausible mechanism of recoordination. Although the interaction between these two forms of attention has so far not been explored in the literature, our modeling provides a first step toward understanding their interplay. We argue that our approach is a parsimonious one. By focusing on dimensional attention, we make use of a preexisting computational principle of category learning models. The theoretical upshot is that we were able to identify a novel mechanism of knowledge restructuring — recoordination driven by a shift in dimensional attention — to supplement the error-driven mechanism that likely underpins relearning (Kalish et al., 2005). In particular, we showed that shifts in dimensional attention, via the exemplar-based gating of competing modules, can lead to extensive behavioral shifts in the absence of further learning. Exemplar-mediated selection of partial knowledge is a defining attribute of ATRIUM and is further supported by our modeling.

Our results complement a number of existing findings suggesting that people rely on exemplar similarity to gate selection of an appropriate subset of task knowledge: Exemplar similarity can drive selection of a rule (Aha & Goldstone, 1992; Erickson & Kruschke, 2002a; see also, Blair, Watson, Walshe, & Maj, 2009), or it can override rule-based responding (Allen & Brooks, 1991). The relation between exemplar similarity and selection of a mode of responding highlights the tight relationship between dimensional and representational attention. Our results suggest that the interaction between these two forms of attention underpins recoordination, which implies separability between task knowledge and the overarching strategy that governs selection

[7]It is an open question whether hints exclusively affect the distribution of dimensional attention or if they also engage alternative mechanisms. For example, Noelle and Cottrell (1995; 1996) modeled the effects of instructions on learning in a recurrent network via dynamic changes in activation state. (We thank an anonymous reviewer for directing our attention to this work.)

and application of that knowledge. Specifically, our results suggest that useful task knowledge (i.e., that which contributes to accurate performance) is actually independent of overarching strategy. If a certain partial knowledge element is particularly useful, it is likely to be incorporated into a number of different strategies. Our modeling with ATRIUM suggests that the distribution of dimensional attention is a principal driver of the selection process. The fact that strategy selection (at least in ATRIUM) is grounded in an underlying exemplar representation leads to a possible theoretical unification of error-free (e.g., recoordination) and error-driven (e.g., relearning) forms of knowledge restructuring.

8.2. Toward a Unified Understanding of Knowledge Restructuring

Our results extend previous work on knowledge restructuring, providing the first direct empirical evidence for recoordination in the absence of relevant pre-experimental knowledge (cf. Little et al., 2006). Further, we demonstrated that recoordination is not limited to cases involving a shift from a more complex strategy to a simpler one. Restructuring was both immediate and symmetrical in both Experiments 1 and 2. We also showed that not all cases of knowledge restructuring are dependent on error (cf. Kalish et al., 2005; see also Dixon & Bangert, 2002; Dixon & Dohn, 2003). People voluntarily restructured their knowledge without requiring additional training, and in Experiment 2, training performance was at ceiling (and essentially error-free for many participants). The results therefore speak to the generality of the recoordination process.

Our findings also extend work on knowledge partitioning (e.g., Kalish et al., 2004; Lewandowsky et al., 2002; Lewandowsky & Kirsner, 2000; Lewandowsky et al., 2006; Little & Lewandowsky, 2009; Yang & Lewandowsky, 2003, 2004). Rather than impeding or even preventing restructuring to an alternative strategy, our results show that knowledge partitioning (i.e., the decomposition of a complex task into modular partial knowledge elements) may actually facilitate strategy change. A unitary knowledge structure (such as ALCOVE's) can only restructure through relearning, which places severe limits on the number of available strategies. By contrast, a modular knowledge structure permits exploration of multiple strategies through the iterative and ad hoc generation of candidate strategies via flexible recoordination or permutation of partial knowledge elements. This notion gels with theoretical perspectives that incorporate some form of hypothesis testing which has received support in category learning (Nosofsky et al., 1994; Nosofsky &

Palmeri, 1998; Rehder & Hoffman, 2005a,b) and elsewhere (Dixon & Kelley, 2007; Rieskamp & Otto, 2006; Siegler, 1996, 1999). It may be the case that recoordination is central to exploration of novel strategic possibilities. We believe this to be a very interesting field of further inquiry.

Although our results highlight aspects of knowledge restructuring that permit retention of different strategies (viz. recoordination), the role of relearning cannot be underestimated. We view recoordination and relearning as complementary processes of knowledge restructuring (cf. Dixon & Kelley, 2007). Recoordination is likely to play an important role in task environments (such as ours) that support several equally valid strategies. Although such environments may actually be quite commonplace (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999, identify a number of "real-world" task environments in which several strategies permit near-ceiling level performance), there is also an abundance of evidence that strategy change is due to relearning involving the partial (or complete) replacement of old strategies with new ones. Such cases have been well-documented in the developmental literature and are associated with shifts from highly error-prone strategies to more sophisticated and accurate ones (e.g., Carey, 1991, 2009; Gopnik & Meltzoff, 1997; Johnson & Carey, 1998; Keil, 1989; Richards & Siegler, 1986; Stafylidou & Vosniadou, 2004; Vosniadou, 1994; Vosniadou & Brewer, 1992, 1994).

A modular, mixture-of-experts view of knowledge is suitably general to accommodate both recoordination and relearning processes (e.g., Appendix A shows that ATRIUM is capable of both). Within this framework, the development of strategic knowledge can be viewed in terms of improving the extent to which modularized partial knowledge elements "match" the task environment. In the search for a strategy that will optimize performance, the process of relearning serves to improve the extent to which partial knowledge contributes to accurate task performance. By contrast, the process of recoordination permits the exploration of different strategies (some of which may not be successful) whilst protecting useful partial knowledge elements from being lost in the process. Our modeling points to a crucial role of dimensional (and thence representational) attention in facilitating this exploration. A rapid attention-shifting mechanism, such as that developed by Kruschke and Johansen (1999) might, in conjunction with a mixture-of-experts architecture, provide a way of instantiating an hypothesis testing mechanism in a connectionist learning model. Kruschke (2001) discussed how such a model might be implemented in a slightly different theoretical context. Ultimately, it is the interaction between recoordination and relearning processes that will determine whether old strategies are replaced by or retained alongside new strategies.

9. Conclusions

We examined the recoordination process of knowledge restructuring in a category learning task. A crucial signature of a recoordination process is that people will continue to have access to old categorization strategies even after having shifted to an alternative. Consistent with recoordination, we showed that when two equally valid strategies are available, knowledge restructuring does not entail loss of old strategies. Instead, people appear to retain and readily recover old strategies in response to a hint. ATRIUM, a mixture-of-experts model, handled the key results via a change in the distribution of dimensional attention, which in turn influenced how the model coordinated its representational modules. The retention of old strategies was possible because learning within ATRIUM occurred at the level of partial knowledge, and was thus insensitive to overarching response strategy. This latter feature of learning is central to recoordination.

A. Appendix A: Details of Category Reversal Shift Simulations

Because the Type 5 problem is readily learned by ATRIUM's exemplar module, we had to "encourage" ATRIUM to learn a rule-plus-exception solution. As such, the models encountered slightly different training regimes. ALCOVE was trained on the Type 5 problem by presenting it with 20 stimulus blocks comprising one presentation of each stimulus in a random order. To encourage rule use in ATRIUM, we modified this training procedure in two ways following relevant precedent (Denton et al., 2008; Erickson & Kruschke, 2002b). First, we included a set of extrapolation items during parameter estimation for ATRIUM. Extrapolation stimuli were defined by extending the category structure by 1 unit in both directions along the x dimension only. For example, Category A extrapolation items were constructed by replicating Items 1, 2, 3, and 4, but giving them an x value of -1 instead of 0. Second, we withheld exception items from the first 8 blocks of training to facilitate rule-based learning (i.e., Items 4 and 8 were withheld for these blocks). This was followed by 12 blocks of training that included all stimuli. An unanticipated side-effect of this training regime was that ATRIUM categorized the

Table A-1: Predicted probabilities of a Category A response for ALCOVE and ATRIUM on the Type 5 problem (Initial), and after the two reversal shifts (Complete and Exception).

		Stimulus							
		1	2	3	4	5	6	7	8
Initial	ALCOVE	1	1	1	0	0	0	0	1
	ATRIUM	1	1	1	.08	0	0	0	.92
	Target	A	A	A	B	B	B	B	A
Complete	ALCOVE	0	0	0	1	1	1	1	0
	ATRIUM	0	0	0	.66	1	1	1	.34
	Target	B	B	B	A	A	A	A	B
Exception	ALCOVE	1	1	1	1	0	0	0	0
	ATRIUM	1	1	1	.86	0	0	0	.15
	Target	A	A	A	A	B	B	B	B

exception stimuli less reliably than ALCOVE (see "Initial" section of Table A-1. The difference in absolute performance between the models is not critical. What is important is that both models were quite successful at learning the Type 5 structure.

For both models, we simulated 25 different training sequences (i.e., 25 simulated learners). Subject to the above constraints, presentation sequences varied randomly across simulated learners. We estimated model parameters by minimizing the root mean squared deviation (RMSD) between model predictions averaged across simulated learners and target category assignments for each stimulus in Table A-1. ATRIUM was equipped with a single rule module orthogonal to the x dimension (centered on $x = 0.5$), and an exemplar module. Parameters used in the Type 5 simulations are reported in Table A-2.

For the reversal shifts, the initial learned weight matrices for the models were pre-loaded, and the same presentation sequences were used (with appropriate changes to category feedback). ALCOVE's weight matrix at the end of each simulation is presented in Table A-3. It is clear that the pattern of changes in the exemplar-to-association weights track the changes in reinforce-

Table A-2: ALCOVE and ATRIUM model parameters for the Type 5 simulations.

	ALCOVE	ATRIUM
c	18.90	7.93
ϕ	29.94	16.00
λ_e	.07	.02
λ_α	.01	.0001
ϕ_g	-	6.48
λ_r	-	1.07
λ_g	-	4.12
γ	-	.87

ment for specific training items. The relative association strengths between category responses reverse, but only if feedback associated with that was altered under a category reversal shift, which is consistent with a relearning process of strategy change.

The weight matrix for ATRIUM's rule module and the exemplar-to-category association weights (for exception items only) are presented in Table A-5. We also report the probability with which the rule module was used to categorize each stimulus in Table A-4, which provides a way of tracking changes in the behavior of the gating mechanism.

In the Complete Reversal simulation, changes in ATRIUM's weight matrices track the relevant changes in reinforcement for individual items, demonstrating a relearning process. In the exception simulation though, the relative ordering of the exception item exemplar-to-category association weights is unchanged relative to their initially learned values. Adaptation in the Exception Reversal simulation was achieved at the level of the gating mechanism (see Table A-4), which shows that the exception items were predominantly categorized by the rule module, rather than the exemplar module, which demonstrates a recoordination process of strategy change.

Table A-3: Weight matrix of ALCOVE for each simulation. Cells denote association weights between category responses and stored exemplars.

		Stimulus							
Simulation	Category	1	2	3	4	5	6	7	8
Initial	A	.75	.75	.75	-.75	-.75	-.75	-.75	.75
	B	-.75	-.75	-.75	.75	.75	.75	.75	-.75
Complete	A	-.57	-.57	-.57	.56	.57	.57	.57	-.56
	B	.57	.57	.57	-.56	-.57	-.57	-.57	.56
Exception	A	.94	.94	.94	.59	-.94	-.94	-.94	-.59
	B	-.94	-.94	-.94	-.59	.94	.94	.94	.59

Table A-4: Probability with which ATRIUM's rule module categorized each stimulus in each simulation.

		Stimulus							
		1	2	3	4	5	6	7	8
P(Rule)	Initial	1	1	1	0	1	1	1	0
	Complete	1	1	1	0	1	1	1	0
	Exception	1	1	1	.56	1	1	1	.52

Table A-5: Weight matrices of ATRIUM for each simulation. Learned weights between large and small valued x dimension inputs and category responses are described in the Lg Rule and Sm Rule columns respectively. Association weights between exemplars representing exception items 4 and 8 and category responses are reported in the Ex4 and Ex8 columns respectively.

Simulation	Category	Lg Rule	Sm Rule	Ex4	Ex8
Initial	A	-1.72	2.72	.01	.19
	B	2.72	-1.72	.19	.01
Complete	A	2.68	-1.68	.20	.16
	B	-1.68	2.68	.16	.20
Exception	A	-1.82	2.82	.16	.17
	B	2.82	-1.82	.17	.16

B. Appendix B: Multidimensional Scaling Study

We assessed the psychological dimensionality of the stimuli used in Experiment 1 via a multidimensional scaling (MDS) study involving a sample of 20 stimuli from the category space used in Experiment 1 (see 3. Stimuli were arranged in a 4×5 grid. Physical coordinates for each stimulus along with the corresponding coordinates returned by the two-dimensional scaling solution are shown in Table B-1.

B.1. Method

B.1.1. Participants

Twenty undergraduate participants from the University of Western Australia completed the MDS study, which was run as a filler task for an unrelated experiment. The experiment lasted under 20 minutes.

B.1.2. Procedure

Participants were instructed to rate the pairwise similarity of a set of stimuli on a scale from 1 (least similar) to 9 (most similar); participants were encouraged to use the full extent of the scale. In line with the categorization experiments, instructions were accompanied by a labeled diagram showing a sample stimulus with arrows pointing to the intended stimulus dimensions: The diameters of the inner and outer circles. Participants were instructed to carefully consider their responses, but without taking more than 1-2 seconds to respond on any given trial.

Prior to collecting ratings, participants previewed the full set of stimuli, presented singly in random order for 1400 ms with a 200 ms blank inter-stimulus interval. Participants then completed the similarity rating task which involved $\binom{20}{2} = 190$ unique stimulus pairs. On each trial, pair members were presented sequentially (in random order) for 1400 ms each (separated by 200 ms). Participants then entered their rating response. Trials were separated by a 600 ms blank screen.

B.2. Results

Similarity ratings were averaged across participants and were submitted to Matlab's MDS routine. The routine minimized the stress (Kruskal, 1964) between the city-block distances between stimuli and proximities that were a monotone function of the similarity ratings. Stress was .2487, .1339, .1015 for the 1-D, 2-D, and 3-D solutions, respectively, suggesting that a two-dimensional representation of the stimuli was most appropriate. The 2D MDS solution as it relates to the physical positioning of the stimuli is presented in Table B-1.

Table B-1: Coordinates of stimuli used in the multidimensional scaling study. Rows define stimuli in terms of the diameter of the outer circle component of the stimulus; columns define the diameter of the inner circle component. Each cell in the table maps a set of physical stimulus coordinates to the corresponding coordinates returned by the two-dimensional multidimensional scaling solution.

	Inner Dimension				
Outer Dimension	120	160	200	240	280
750	-2.47, 1.80	-1.68, 1.75	-0.56, 1.62	0.24, 1.60	0.93, 1.35
650	-2.16, 1.08	-1.36, 0.90	-0.08, 0.84	0.90, 0.66	1.50, 0.43
550	-2.01, -0.29	-0.97, -0.40	0.58, -0.43	1.31, -0.63	1.96, -1.03
450	-1.45, -1.47	-0.17, -1.49	0.98, -1.71	1.92, -1.91	2.58, -2.68

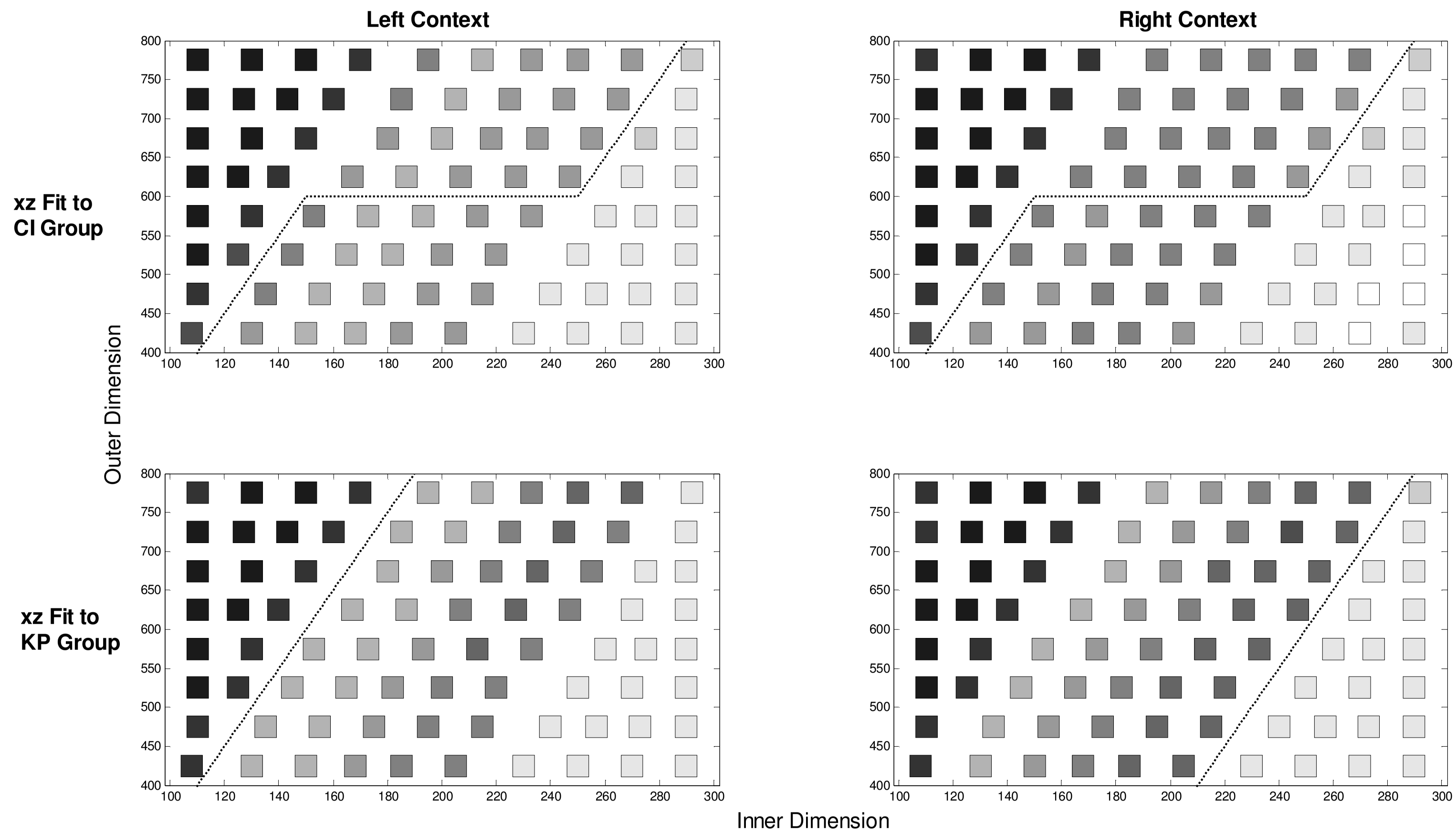


Figure B-1: ATRIUM fit to data from transfer test 1 performance of the Initial-CI and Initial-KP groups from Experiment 1 using x and z dimensions as input to the exemplar module.

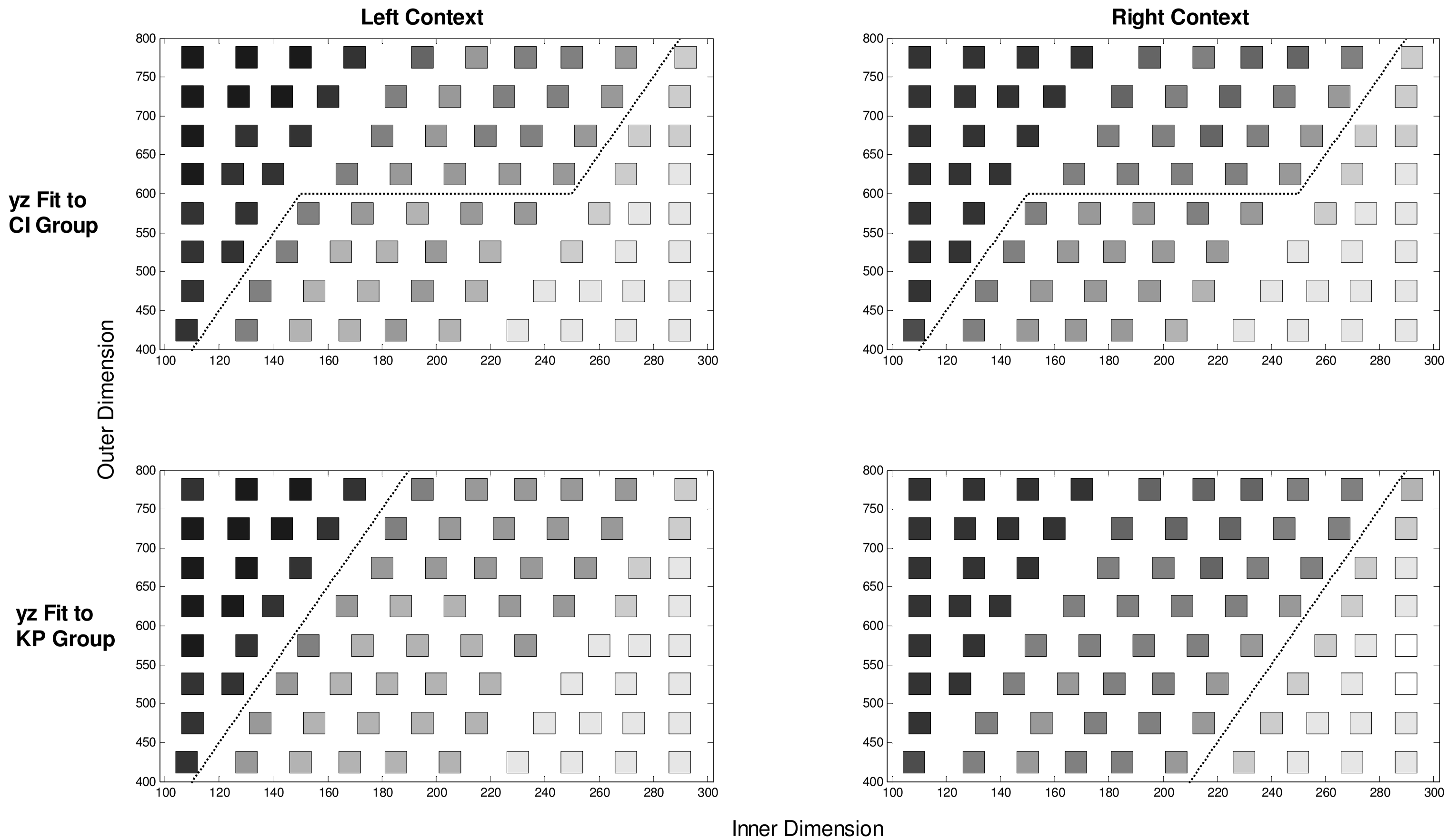


Figure B-2: ATRIUM fit to data from transfer test 1 performance of the Initial-CI and Initial-KP groups from Experiment 1 using y and z dimensions as input to the exemplar module.

Although the MDS solution was clearly two-dimensional, its axes did not fully align with the physical dimensions of the stimuli, raising the possibility that people may have relied on psychological dimensions other than those used in the category space. Specifically, instead of perceiving stimuli solely with respect to the diameters of the inner and outer circles (denoted x and y, respectively), people may have created a compound dimension (which we call z) that represents the thickness of the ring formed by the two component circles.

Because the physical structures that correspond to the xy, xz, and yz solutions are approximate rotations of one another, we sought to identify the appropriate space via modeling. We fit ATRIUM to the Initial-KP group and Initial-CI group data from Experiment 1, representing stimuli in terms of the context dimension and one of the three potential 2-D configurations. In all simulations, the left and right partial boundaries were represented along the z dimension. Only the dimensional inputs to the exemplar module differed across simulations. ATRIUM was only able to model the data given the context and xy configuration. Under both the xz and yz configurations, the model's behavior deviated considerably from the behavior of human participants (most notably when fitting the Initial-KP group), respectively generating noisy striped (xz) and sawtooth (yz) response profiles (see Figures B-1 and B-2). This suggests that participants represented the stimuli along the physical stimulus dimensions (i.e., according to the diameters of the inner and outer component), and gels with the protocols collected from participants following completion of the experiments (see main text for details).

References

Aha, D. W. & Goldstone, R. L. (1992). Concept learning and flexible weighting. In J. K. Kruschke (Ed.), *Proceedings of the 14th Annual Conference of the Cognitive Science Society* (pp. 534–539). Hillsdale, NJ: Erlbaum.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716–723.

Allen, S. W. & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, 120, 3–19.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98, 409–429.

Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 14, 33–53.

Ashby, F. G. & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93, 154–179.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, 11, 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 11, 629–654.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 35, 1196–1206.

Bourne, L. E., Healy, A. F., Kole, J. A., & Graham, S. M. (2006). Strategy shifts in classification skill acquisition: Does memory retrieval dominate rule use? *Memory & Cognition*, 34, 903–913.

Bourne, L. E., Healy, A. F., Parker, J. T., & Rickard, T. C. (1999). The strategic basis of performance in binary classification tasks: Strategy choices and strategy transitions. *Journal of Memory & Language*, 41, 223–252.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: Essays in biology and cognition.* (pp. 257–291). Hillsdale, NJ: Erlbaum.

Carey, S. (2009). *The origin of concepts*. Oxford: Oxford University Press.

Chase, W. G. & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York: Academic Press.

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

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

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science*, 9, 1–7.

Denton, S. E., Kruschke, J. K., & Erickson, M. A. (2008). Rule-based extrapolation: A continuing challenge for exemplar models. *Psychonomic Bulletin & Review*, 15, 780–786.

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age: The Jean Piaget symposium series.* (pp. 49–70). Hillsdale, NJ: Erlbaum.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science*, 28, 843–900.

Dixon, J. A. & Bangert, A. S. (2002). The prehistory of discovery: Precursors of representational change in solving gear system problems. *Developmental Psychology*, 38, 918–933.

Dixon, J. A. & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition*, 31, 1082–1093.

Dixon, J. A. & Kelley, E. (2007). Theory revision and redescription: Complementary processes in knowledge acquisition. *Current Directions in Psychological Science*, 16, 111–115.

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, 36, 749–761.

Erickson, M. A. & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107–140.

Erickson, M. A. & Kruschke, J. K. (2002a). *Multiple representations in inductive category learning: Evidence of stimulus- and task-dependent representation.* Unpublished manuscript.

Erickson, M. A. & Kruschke, J. K. (2002b). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin & Review*, 9, 160–168.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple heuristics that make us smart.* New York, NY: Oxford University Press.

Gopnik, A. & Meltzoff, A. N. (1997). *Words, thoughts, and theories.* Cambridge, MA: MIT Press.

Hartnett, P. & Gelman, R. (1998). Early understandings of numbers: Paths or barriers to the construction of new understandings? *Learning & Instruction*, 8, 341–374.

Jacobs, R. A. (1999). Computational studies of the development of functionally specialized neural modules. *Trends in Cognitive Sciences*, 3, 31–38.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3, 79–87.

Johansen, M. K. & Palmeri, T. J. (2002). Are there representational shifts during category learning? *Cognitive Psychology*, 45, 482–553.

Johnson, S. & Carey, S. (1998). Knowledge enrichment and conceptual change in folkbiology: Evidence from Williams syndrome. *Cognitive Psychology*, 37, 156–200.

Kalish, M. L., Lewandowsky, S., & Davies, M. (2005). Error-driven knowledge restructuring in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 846–861.

Kalish, M. L., Lewandowsky, S., & Kruschke, J. K. (2004). Population of linear experts: Knowledge partitioning and function learning. *Psychological Review*, 111, 1072–1099.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

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

Kruschke, J. K. (1996). Dimensional relevance shifts in category learning. *Connection Science*, 8, 225–247.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology*, 45, 812–863.

Kruschke, J. K. (2008). Models of categorization. In R. Sun (Ed.), *The Cambridge Handbook of Computational Psychology* (pp. 267–301). Cambridge: Cambridge University Press.

Kruschke, J. K. & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 25, 1083–1119.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29, 115–123.

Lewandowsky, S., Kalish, M., & Griffiths, T. L. (2000). Competing strategies in categorization: Expediency and resistance to knowledge restructuring. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26, 1666–1684.

Lewandowsky, S., Kalish, M., & Ngang, S. K. (2002). Simplified learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General*, 131, 163–193.

Lewandowsky, S. & Kirsner, K. (2000). Knowledge partitioning: Context-dependent use of expertise. *Memory & Cognition*, 28, 295–305.

Lewandowsky, S., Little, D. R., & Kalish, M. L. (2007). Knowledge and expertise. In F. T. Durso (Ed.), *Handbook of applied cognition.* (pp. 83–110). Hoboken, NJ: Wiley.

Lewandowsky, S., Roberts, L., & Yang, L.-X. (2006). Knowledge partitioning in categorization: Boundary conditions. *Memory & Cognition*, 34, 1676–1688.

Little, D. R. & Lewandowsky, S. (2009). Beyond non-utilization: Irrelevant cues can gate learning in probabilistic categorization. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 530–550.

Little, D. R., Lewandowsky, S., & Heit, E. (2006). Ad hoc category restructuring. *Memory & Cognition*, 34, 1398–1413.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309–332.

Macho, S. (1997). Effect of relevance shifts in category acquisition: A test of neural networks. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 23, 30–53.

Medin, D. L. & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.

Medin, D. L. & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning & Memory*, 7, 355–368.

Nelder, J. A. & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–313.

Noelle, D. C. & Cottrell, G. W. (1995). A connectionist model of instruction following. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th Annual Cognitive Science Conference* (pp. 369–374). Hillsdale, NJ: Erlbaum.

Noelle, D. C. & Cottrell, G. W. (1996). Modeling interference effects in instructed category learning. In G. W. Cottrell (Ed.), *Proceedings of the 18th Annual Cognitive Science Conference* (pp. 475–480). Hillsdale, NJ: Erlbaum.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 10, 104–114.

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 13, 87–108.

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 15, 282–304.

Nosofsky, R. M. & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, 7, 375–402.

Nosofsky, R. M. & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, 5, 345–369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, 101, 53–79.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.

Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.

Rehder, B. & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning. *Cognitive Psychology*, 51, 1–41.

Rehder, B. & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and models of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 811–829.

Richards, D. D. & Siegler, R. S. (1986). Children's understandings of the attributes of life. *Journal of Experimental Child Psychology*, 42, 1–22.

Rickard, T. C. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General*, 126, 288–311.

Rieskamp, J. & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.

Rodrigues, P. M. & Murre, J. M. J. (2007). Rule-plus-exception tasks: A problem for exemplar models? *Psychonomic Bulletin & Review*, 14, 640–646.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–233.

Shafto, P. & Coley, J. D. (2003). Development of categorization and reasoning in the natural world: Novices to experts, naïve similarity to ecological knowledge. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 29, 641–649.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General & Applied*, 75, 1–42.

Shrager, J. & Siegler, R. S. (1998). A model of children's strategy choices and strategy discoveries. *Psychological Science*, 9, 405–410.

Siegler, R. S. (1981). Developmental sequences within and between concepts. *Monographs of the Society for Research in Child Development*, 46, 1–74.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, 116, 250–264.

Siegler, R. S. (1996). *Emerging minds: The process of change in children's thinking*. Oxford: Oxford University Press.

Siegler, R. S. (1999). Strategic development. *Trends in Cognitive Sciences*, 3, 430–435.

Siegler, R. S. & Jenkins, E. (1989). *How children discover new strategies.* Hillsdale, NJ: Erlbaum.

Smith, J. D. & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 24, 1411–1436.

Stafylidou, S. & Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learning & Instruction*, 14, 503–518.

Staszewski, J. J. (1988). Skilled memory and expert mental calculation. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 71–128). Hillsdale, NJ: Erlbaum.

Stephen, D. G., Dixon, J. A., & Isenhower, R. W. (2009). Dynamics of representational change: Entropy, action, and cognition. *Journal of Experimental Psychology: Human Perception & Performance*, 35, 1811–1832.

Vosniadou, S. (1994). Universal and culture-specific properties of children's mental models of the earth. In L. Hirschfeld & S. A. Gelman (Eds.), *Mapping the mind: Domain specificity in cognition and culture.* (pp. 412–430). Cambridge, MA: Cambridge University Press.

Vosniadou, S. & Brewer, W. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24, 535–585.

Vosniadou, S. & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science*, 18, 123–183.

Yang, L.-X. & Lewandowsky, S. (2003). Context-gated knowledge partitioning in categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 29, 663–679.

Yang, L.-X. & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30, 1045–1064.