

Journal of Experimental Psychology: Learning, Memory, and Cognition

Sorting Out the Problem of Inert Knowledge: Category Construction to Promote Spontaneous Transfer

Kenneth J. Kurtz and Garrett Honke

Online First Publication, July 25, 2019. <http://dx.doi.org/10.1037/xlm0000750>

CITATION

Kurtz, K. J., & Honke, G. (2019, July 25). Sorting Out the Problem of Inert Knowledge: Category Construction to Promote Spontaneous Transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. <http://dx.doi.org/10.1037/xlm0000750>

Sorting Out the Problem of Inert Knowledge: Category Construction to Promote Spontaneous Transfer

Kenneth J. Kurtz and Garrett Honke
Binghamton University, SUNY

A fundamental goal in the study of human cognition is to understand the transfer of knowledge. This goes hand-in-hand with the translational goal of promoting such transfer via instructional techniques. Despite a rich history of research using the analogical problem-solving paradigm, no study activity has been found to produce a robust rate of successful spontaneous transfer—even when the test is immediate. We propose the *category status hypothesis* as an explanation of the difficulty of transfer and as motivation for a novel approach to promoting transfer. We report a set of experiments evaluating a category construction technique based on a sorting task. In Experiment 1a, we found category construction to be significantly more effective than the “gold standard” of schema abstraction through comparison of 2 analogous cases. In Experiment 1b, we explored a variation of the category construction technique that did not reliably differ in effectiveness from comparison-based schema abstraction—we also verified that both study tasks were superior to a baseline task of separate summarization of 2 cases. In Experiment 2, we conducted a replication of the initial design with higher power and confirmed the significant advantage for category construction over schema abstraction via comparison. In Experiment 3, we compared category construction to an information-consistent reading comprehension control to evaluate competing interpretations of the category construction advantage found in Experiments 1 and 2. We discuss theoretical and applied implications of these findings.

Keywords: knowledge transfer, learning, relational categories, inert knowledge

Supplemental materials: <http://dx.doi.org/10.1037/xlm0000750.supp>

Much of the importance of learning is building toward future success, that is, changing what a person is capable of going forward. In this sense, the true test of knowledge gained is whether that knowledge can be accessed and applied when it pertains to new problems and situations that arise. In the most demanding case: Can the right knowledge be brought to bear when there are no or few cues to its relevance? Understanding what underlies the fluid access and application of acquired knowledge in problem solving and reasoning is a central goal for researchers and educators alike (National Research Council, 2012). While the goal is clear, little agreement is found on key issues such as how often

transfer occurs (e.g., Day & Goldstone, 2011) or how to characterize transfer situations (Barnett & Ceci, 2002). Despite this lack of agreement, transfer is frequently used as a measure of students' mastery of a topic and as a measure of success for education programs generally (Engle, 2012).

One area of agreement is that relational information is critical to cross-domain transfer (Gentner, 1983). This is because relational content is what is likely to remain consistent across problems for which a common principle can be applied. Domain experts use structural correspondences, matching relations between elements (objects, features, contexts, etc.), as the basis for organizing their knowledge and recognizing similarity among problems, while novices instead tend to focus on surface elements (Chi, Feltovich, & Glaser, 1981; Rottman, Gentner, & Goldwater, 2012). Is it possible to make every day thinking and problem-solving more structurally sensitive, that is, more “expert-like,” but without the long road and domain limitation? The goal of the present project is to theoretically motivate and empirically evaluate a category construction technique for building knowledge representations that are portable and can be spontaneously accessed from a cue that is relationally similar, but lacks superficial similarity.

Inert Knowledge

The basis for traditional perspectives on transfer extends back to Thorndike's (1924) notion that shared elements between situations are required for knowledge transfer. The dominant view of transfer in analogical problem solving research is that a process of

Kenneth J. Kurtz and Garrett Honke, Department of Psychology, Binghamton University, SUNY.

Portions of this article are adapted from Garrett Honke's Master's thesis. A preprint version is hosted on PsyArXiv. All materials, data, and analysis are hosted on OSF (<https://osf.io/xc8qm/>). This article may not exactly replicate the final version published in the APA journal. It is not the copy of record. We thank Katy Kam for her contributions in the development of this project, Sarah Laszlo and Peter Gerhardstein for comments on a draft of the document, and the members of the Learning and Representation in Cognition (LARC) Laboratory at Binghamton University.

Correspondence concerning this article should be addressed to Kenneth J. Kurtz, Department of Psychology, Binghamton University, SUNY, P.O. Box 6000, Binghamton University, Binghamton, NY 13902-6000. E-mail: kkurtz@binghamton.edu

similarity-based retrieval activates knowledge of previous experience. Most such reminders are mundane or limited in their utility, but the mechanism includes the potential to produce analogical reminders—when relevant knowledge is accessed that is not literally similar to the target. This is based on a similarity match in terms of underlying structural content with little or no match in terms of surface qualities. Over several decades, research has focused on efforts to develop study techniques that raise the likelihood of retrieving relationally matched knowledge (for review, see Day & Goldstone, 2012).

Laboratory studies often use a two-stage (A-B) transfer paradigm in which participants receive: (A) one or more cases to evaluate under a study task, and then (B) a new problem that (unbeknownst to the participant) can be solved using the relational principle presented in (A). Findings based on this research show that people struggle—they typically fail to access stored analogs (even those just encountered) to solve new problems if the cue and source cases lack surface similarity (Bassok & Holyoak, 1989; Gentner, Rattermann, & Forbus, 1993; Gick & Holyoak, 1980, 1983; Gonzalez & Wong, 2012; Holyoak & Koh, 1987; Novick, 1988; Perkins, 2009; Reeves & Weisberg, 1994; Ross, 1987; Spencer & Weisberg, 1986; Trench & Minervino, 2015). This broad-based failure to access the relevant content from memory is referred to as the inert knowledge problem (Whitehead, 1929). The knowledge is encoded and retrievable when directly cued (Gick & Holyoak, 1980), but spontaneous reminding of a concept or situation that shares relational similarity is relatively rare. This is a puzzling pattern given that people consider relationally similar cases to comprise more sound analogies (Gentner et al., 1993)—people grasp and appreciate relational similarity, but rarely does it guide reminding.

The status quo account appeals to characteristics of similarity-based memory retrieval, that is, the degree of match between the cue and what is stored in memory. Forbus, Gentner, and Law (1995) argue that a structurally sensitive search of memory is too computationally demanding and, therefore, requires an initial “dumb” search; accordingly, because most of the overall content consists of objects and attributes, the preponderance of what is retrieved is based on superficial match. In an influential report, Gick and Holyoak (1983) found that guided, effortful comparison of two copresented analogs improves the likelihood of subsequent spontaneous retrieval and transfer. Research in domains as diverse as mathematics (Rittle-Johnson & Star, 2007, 2009), science (Goldstone & Son, 2005; Kurtz, Miao, & Gentner, 2001), engineering (Gentner et al., 2016), statistics (Quilici & Mayer, 1996), negotiation (Gentner, Loewenstein, & Thompson, 2003), and anatomy (Kurtz & Gentner, 2013) has strengthened the position that experiencing an analogy between sources facilitates the learning and transfer of relational content (see Alfieri, Nokes-Malach, & Schunn, 2013, for review and meta-analysis). When comparison highlights shared relational content, this promotes the abstraction of shared relational structure (Gentner, 1989, 2010) in what is often referred to as a schema. A widely held hypothesis is that schema abstraction promotes transfer by increasing the similarity between a retrieval cue and abstract stored knowledge (Gentner et al., 1993; Gick & Holyoak, 1983; Vendetti, Wu, Rowshanshad, Knowlton, & Holyoak, 2014; see also Loewenstein, 2010). According to this *memory* account, structurally consistent sources are more accessible in abstracted form because the schema retains

relational similarity to an analogous cue without the baggage of superficial elements that lead to mismatches and restrict access. As such, study techniques guided by the memory account aim to make knowledge more accessible to the learner by promoting the induction of more abstract representations (usually via guided comparison of two source analogs).

Category Status Hypothesis

Over 30 years ago, comparison-based learning was shown to produce a significant boost in transfer rates while nothing else proved effective (Gick & Holyoak, 1983). Under the easiest conditions for spontaneous transfer, a guided comparison task at study typically leads to problem-solving success for one-quarter to one half of participants. This raises the question: Is it really so hard for people to notice that a new problem is fundamentally like the ones they just saw? Is there a fundamental limitation intrinsic to the cognitive system—or are researchers somehow not thinking about this right? Unlike the sociomotivational critique of transfer research (e.g., Greeno, Smith, & Moore, 1993; Lave, 1988; Lobato, 2012), we maintain the basic perspective from the memory account that the inert knowledge problem can be addressed through an information-processing approach directed at promoting more portable representations. We also maintain a commitment to the role of abstraction. The key difference is that we propose viewing the underlying information processing not just in terms of the memory account (i.e., the mechanics of storage and retrieval), but also in terms of another cognitive process: categorization.

How is categorizing different than reminding as a path from cue to stored knowledge? It is more directed and more active. There is a specific goal: to activate/select the best conceptual knowledge with which to recognize and make sense of current experience. The selection of an appropriate category is a search for an *understanding* of the target—a complex process, yet one that people engage in naturally and often automatically. In addition to being fluidly and intrinsically activated, the effort to make sense of a target in terms of prior knowledge is a direct psychological priority (addressing the questions: What is it and what does it mean to me?), as opposed to the “ocean of possibilities” that can be retrieved through association-based reminding (Loewenstein, 2010). Categorization launches a constellation of powerful and systematic cognitive activity: construal of the stimulus as a category member, generation of inferences and predictions, updating of category knowledge, activation of language, and semantic memory (Murphy, 2002).

How does categorization relate to analogical transfer? We suggest that an individual faced with a problem may implicitly compute the answer not only to “What does this remind me of?” but also to: “What kind of situation is this?” because the standard way people make sense of the world is by fitting what they experience to a known category. There is an extensive history of research devoted to understanding attribute-based categorization (e.g., Bruner, Goodnow, & Austin, 1956; Komatsu, 1992; Kurtz, 2015; Murphy, 2002; E. E. Smith & Medin, 1981), but less work on relational categories that cohere around a common relational quality held by all members (Gentner & Kurtz, 2005; Markman & Stilwell, 2001). The learning and use of relational categories, however, has become the subject of increased research attention (e.g., Corral & Jones, 2014; Corral, Kurtz, & Jones, 2018; Gentner,

2005; Goldwater, Markman, & Stilwell, 2011; Higgins, 2017; Kurtz, Boukrina, & Gentner, 2013; Patterson & Kurtz, 2015, 2016; Rehder & Ross, 2001).

People rarely spontaneously connect their immediate experience to relationally consistent knowledge without elemental (object) matches. A new explanation is that candidate sources for analogical reminding are not functioning as categories and, therefore, a categorization-driven path (“What kind of situation is this?”) to accessing relevant knowledge comes up empty. To remedy this, abstraction from individual cases is required, but not in the form of a filter that removes the idiosyncratic surface content from individual cases (see also Corral & Jones, 2014). Instead, the abstraction would be the formation of a relational category that is more than just a record in memory of the common content across two analogs. Consider a relational category for situations like this: you observe Callie steal an important opportunity from Ethan and then later notice Ethan laying a trap for Callie. This is an example of “revenge.” The point is that some relational categories are sufficiently robust in their utility that they are fluidly and spontaneously activated to guide the online construal of experience. Can any knowledge become realized as a relational category that is ready at hand to make sense of experience?

This framing in terms of a categorization-based path offers new ways to think about why analogical transfer is hard and how it can be improved. First, the categorization path may draw upon a different knowledge base than the memory path. It may be that simply storing a case or even storing a set of structurally alike cases with an abstract schema is helpful to the memory path, but is unlikely to promote success along the categorization path. This is because more is required for a knowledge representation to be a category, that is, a stable tool of thought used to pick out a particular kind of meaning in the world and to fit that meaning into a network of semantic knowledge. This is the difference between representing one or more cases of experienced content and representing that content as a category. A knowledge representation with category status does explanatory work—it is an ontological and taxonomic commitment: a belief in a set of semantic content that picks out a useful meaning in the world (category intension) and that systematically organizes the world into examples and nonexamples (category extension). With regard to transfer, the idea is that an analogical source can be accessed only through the memory path unless it becomes part of a pool of categorical knowledge, that is, actual concepts as opposed to stripped-down stored cases. A study task that promotes category status for a principle would open the door to categorization-based analogical transfer; that is, noticing that a situation corresponds to a known type. The categorization path means that analogical transfer becomes more likely because: (a) a natural process of making sense of the stimulus is invoked by the cue; (b) the potential to achieve a categorization can arise from reasoning processes such as theory-like explanation rather than similarity (Murphy & Medin, 1985); and (c) even an exhaustive search is more likely to be fruitful given the relatively compact search space of concepts as opposed to that of all experienced cases. This notion of different pools of knowledge for reminding and categorizing is not necessarily a standard view (though one could link it to the classic distinction between semantic and episodic memory systems, e.g., Tulving, 1972).

Gentner and Kurtz (2005) discuss the psychological role and importance of relational categories—including the observation that

relational categories may not generally play an identification role in finding correspondences between immediate experience and prior knowledge. We suggest that practice with the process of mapping back and forth between an abstraction and specific cases could serve as an important factor. It may not be enough simply to have an abstraction—if the goal is for that abstraction to be accessible, then it is important to build a path that can later be more easily traveled between the generic and specific. The process of constructing a concept invested with meaning as a *kind of thing to understand about the world* may depend on a history of functionally successful mappings between concrete experience and abstraction.

Under the view that the success of the categorization-based path to analogical transfer arises from either (a) knowledge entering the pool of ontologically/taxonomically significant categorization options for a classification mechanism or (b) practice strengthening the path to and from the abstraction, we put forward a practical hypothesis. Given that it is relatively straightforward to view the principles or knowledge addressed in studies of analogical transfer as relational categories (see also Goldwater & Schalk, 2016), the *category status hypothesis* states that: to the extent that the form of a knowledge representation is more category-like, the knowledge will be easier to access under the critical conditions of high structural match and low superficial match. On this basis, we suggest that people learn to use relation-based categories in much the same way as attribute-based categories—natural kinds (e.g., birds) and artifacts (e.g., chairs)—to make sense of everyday experience. Key factors that are likely to promote category status include the opportunity to learn inductively from multiple examples and to practice using the knowledge as a category—specifically by invoking the category with regard to specific cases (and not to others) and by invoking specific cases as instantiations of the category. In the present research, we investigate a study technique to encourage category status in this manner.

Learning to Transfer With Category Construction

We propose a category construction technique to encourage the formation of a knowledge representation with increased category status. Historically, category construction (a sorting task) has been used as an assessment measure for existing knowledge (e.g., Chi et al., 1981; see also the ambiguous sorting task, Goldwater & Gentner, 2015; Rottman et al., 2012) and as an unsupervised learning task to study biases and constraints on conceptual organization (e.g., Imai & Garner, 1965, 1968; Medin, Wattenmaker, & Hampson, 1987; Milton & Wills, 2004). We use category construction as an inductive learning technique to drive the discovery of generic knowledge structures by organizing a set of items into classes reflecting within-group similarity and between-groups differentiation (without the benefit of labels or feedback). This task neatly captures our proposal for promoting category status: a set of examples to guide inductive learning, an encouragement to think of the principle in categorical terms, experience with the contrast between members and nonmembers, and experience mapping between a generic form and multiple concrete instances.

We use a version of category construction in which all items are presented for simultaneous consideration and no direction is provided to guide the sort except for a restriction to form two groups. The materials to be sorted are six text passages with half of the

cases conforming to a target principle (relational category). The process of assigning cases to categories is unconstrained, but participants are expected to engage in a combination of assessing individual items, comparing items, and identifying/evaluating potential bases for group membership. The task requires the learner to engage in the goal of assimilating a set of far-ranging examples into a collective entity—this focus on the process of recognizing examples in varied contexts as “going together” addresses the lack of identifiable cohesion between target and source that makes transfer hard. Category construction trains the learner to identify different manifestations of a common principle and to rule out examples that fall outside of this common basis. The learner is called upon not only to induce an accurate and general version of a principle, but also to recognize the range of manifestations of the principle and successfully distinguish nonmembers.

Critical to the specification of a category construction study task is the type of similarity between examples used in the task. First, we consider the similarity of the cases representing the target principle. Prior research suggests that high similarity between examples will increase the ease of alignment and facilitate the recognition of differences in their relational structure (Gentner & Markman, 1995; Goldstone, 1994; Sagi, Gentner, & Lovett, 2012). However, this approach carries a potential cost in that highly similar examples can limit the generalization and transfer of a principle to superficially distinct, but relationally similar problems (Goldstone & Sakamoto, 2003; Sloutsky, Kaminski, & Heckler, 2005). We opted to prioritize the latter consideration and make all of the cases in the sort sets fully distinctive in their surface properties (agents, objects, and settings). In addition, there is the question of the relationship between the cases instantiating the target principle and those in the contrast set. We sought to highlight the target principle as a category in two very different ways: (a) a coherent contrast set organized around a principle related to the target category but with a key alignable difference (Markman & Gentner, 1996) as in the families of related concepts often present in classroom content; and (b) an incoherent contrast set lacking any commonalities with one another or the target set.

Unsolved or Solved Problems as Study Cases

To supplement our goal of advancing basic understanding and developing a study technique to promote the best possible transfer outcomes, we also sought to evaluate the effectiveness of study examples that are problem statements without solutions. To be clear, this means the target principle for transfer is *never* presented either in abstract or contextualized form. Research on the effect of comparison learning and other case-based reasoning approaches has most often included the use of problems with the solution (or

principle) included (e.g., Catrambone, 1996, 1998; Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Nokes & VanLehn, 2008; Rittle-Johnson & Star, 2007, 2009; Scheiter, Gerjets, & Schuh, 2010). Less work has addressed the use of unsolved examples—despite the fact that natural encounters with a problem start (and commonly enough end) without a solution. This is not just a consideration in terms of ecological validity—experience with unsolved examples during study provides transfer-appropriate processing (Morris, Bransford, & Franks, 1977) relative to the task at test. We are not the first to note this behavioral correspondence between generating solutions during study and doing so during test (see Lockhart, Lamon, & Gick, 1988). Along these lines, Needham and Begg (1991) showed that participants who were asked to solve a problem during study transferred the subsequently presented solution to novel problems more often than participants who only studied the problem and then received the solution. Aside from increasing the similarity between study and test conditions, using unsolved examples might also afford an opportunity to benefit from a desirable difficulty in the task (Bjork, 1994) or a type of retrieval practice such as that associated with test-enhanced learning (e.g., Roediger & Karpicke, 2006).

On the other hand, results from classroom-based research on comparison suggest that providing a solution during comparison facilitates learning. Rittle-Johnson and Star (2009) had students learn linear equations by comparing examples that were either (A) equivalent problems with the same solution, (B) distinct problems with the same solution, or (C) different solution methods for the same problem. They found comparing different solutions to the same problem produced the highest observed measures of conceptual knowledge and procedural flexibility on test problems, followed by comparing distinct problems with the same solution. The learning tasks with unsolved examples presented in this research are most similar to the (B) condition (distinct problems with one solution condition) from Rittle-Johnson and Star (2009). A direct comparison of these competing design options for learning materials will provide a better understanding of their relative benefits for learning with comparison and category construction.

Experiment 1a

Our primary goal was to evaluate the effectiveness of category construction, so we elected to assess the sort task relative to the well-established approach of comparison-based schema abstraction (described below). Certain aspects of the category construction task were treated as defaults, rather than candidates for experimental manipulation. Specifically, we chose a low-complexity target principle (using *disguise to harm an adversary*) for two important reasons (see Table 1). One is that our investigation of

Table 1
Study and Transfer Test Passages

Sample study example	Transfer test query
Carnivorous plants eat insects. Insects do not naturally approach these plants; instead the insects are drawn to the scent of rotting animal flesh. Some carnivorous plants all have the ability to emit chemicals into the air that mimic particular smells.	Once a virus enters your body, it infects a cell by injecting DNA into it. This viral DNA instructs the cell to produce thousands of copies of the virus. However, no matter how many virus copies are made the immune system always recognizes them as foreign and attacks when they exit the cell.
How do the plants get their food?	How can the virus leave the cell AND survive?

unsolved problems at study requires it—a complex target principle would not be compatible with the presentation of only unsolved cases because the solution principle would only rarely be discovered. More broadly, the rationale is to address a continual challenge in research on transfer: if a principle is more complex, then performance is necessarily affected by both the degree of mastery and the degree of transfer and it is difficult to decipher what is responsible for what. Has transfer failed because the knowledge was not satisfactorily in place or because of an actual failure to transfer that knowledge? Does a treatment succeed because it promoted transfer or merely because it promoted mastery? Our core question (i.e., the inert knowledge problem) is about transfer of intact, available knowledge. To best address this question we use a low-complexity principle that can be reliably mastered by participants.

To be clear, we have carefully chosen a low-complexity principle that is still sufficiently interesting to conform to the requirements of an analogical problem solving task. In comparative terms, Duncker's Radiation problem includes three required relational constraints (powerful force, obstacle to direct application, division, and simultaneous convergence) compared with the two requirements of the *disguise to harm* principle. Additionally, we note that participants are likely to have prior familiarity with our principle—we take this as a given that does not impact our core question of transfer (and removes the variability inherent in using a target principle with which participants may have varied degrees of prior understanding). Further, we consider these problems as a more challenging test of transfer than problems calling for procedural knowledge transfer (such as the application of an algebra formula in distinct domains) in that the key variables are not explicitly identified and the possible operations extend far beyond those possible for procedural content such as math problems. In summary, it would be a serious problem if our low-complexity target principle did not show the standard phenomenology of rare spontaneous transfer, but (as will be seen) that was not at all the case.

With the benefit of pilot work, we selected a sort set of six cases to achieve a balance between too few and too many examples to evaluate. We opted for minimal instructional guidance to let the sort task make its own demands on the participant—although we did specify sorting into two equal-sized groups to help keep participants on track. After completion of the sort, we asked participants to write a brief explanation of the similarities of the cases they grouped together. No evaluative feedback was provided for any component of the study task.

With these procedural defaults in place, the remaining choices had to do with the nature of the cases to be used in the sort task. As discussed above, we selected a contrast set for this study that cohered around a principle related to the target principle (see Appendices B and C). The use of neighboring relationally coherent categories provides a potential basis for improving the quality of the representation of the target principle. Noticing distinctions in the face of high relational similarity requires close attention to the relational structure of the examples (e.g., Sagi et al., 2012). The other issue regarding the content of the cases is the presentation of problem situations with or without solutions. Using highly straightforward (and likely familiar) principles allowed us to investigate the use of solved versus unsolved cases within the category construction task. To be clear, we expected that the underlying principles could be readily discovered from cases pre-

sented as problems without solutions. In summary, we used a 2×2 between-subjects design to address the influence of study task (comparison-based schema abstraction vs. category construction) and example type (solved vs. unsolved) on analogical problem solving. We predicted that category construction would promote transfer because of the emphasis on category status and we were interested to see how the study techniques would interact with the use of solved versus unsolved examples.

Method

Participants and materials. Undergraduate students from Binghamton University were recruited from the Psychology Department pool and participated for credit toward completion of a course requirement. All participants gave informed consent and the experiment was approved by the Institutional Review Board (IRB) of Binghamton University. Participants ($N = 119$) were randomly assigned to one of four conditions. The experiment was conducted on paper in two consecutive phases (following the A–B paradigm) as part of an hour-long experimental session consisting of a series of unconnected cognitive tasks. The target and contrasting principles, the two cases used for comparison, and the additional positive case plus the three contrast cases used for category construction are provided in the Appendices B and C. In the test phase, participants were given a transfer problem in an unrelated domain that could be solved using the target principle. All study materials, data, and analyses are archived on OSF and also provided as a file of online supplemental materials included with the online publication.¹

Procedure. The procedure closely follows the standard A–B transfer paradigm, where participants are provided with a task and materials during a study phase and then assessed on their ability to use what they may have learned in a testing phase. The study phase in the present experiments was either a category construction task, a comparison-based schema abstraction task, or a summarization task (Experiment 1b only). Aside from being consecutive tasks in the experimental session, the testing phase provided no cues that might alert the participants that the study and test phases were related. At test, participants provided a written response to the transfer problem and then indicated how familiar they were with it (e.g., if they had ever seen it before). At no point during the experiment were participants informed about the critical relational similarity (and, thus, solution strategy) shared between the study and test materials.

Participants who received the category construction study task were given a sorting mat and six cards: three cards containing examples of the target principle and three cards that represented a neighboring principle. Depending upon the condition, the cards contained either all unsolved examples ($n = 30$; Appendix C) or all solved examples ($n = 29$; Appendix B). The instructions were the same:

You are about to read six cards that each demonstrate a different problem situation. These passages can be split into two separate groups of similar problems and solutions. Your job is to group the passages that are solvable using the same general principle. Please

¹ <https://osf.io/xc8qm/>.

consider the passages carefully to make sure you are satisfied with your sort.

After the task was completed, participants were asked to provide a written description of the problem and solution common to each group that they created.

Participants receiving the comparison study task were given a sheet with two horizontally aligned cases that instantiated the target principle (these were the same as two of the positive cases in the sort task). The examples were presented either solved ($n = 30$) or unsolved ($n = 30$). The instructions read: "You are about to read two passages that demonstrate different situations. Please read each passage carefully and then proceed to the following questions." After reading the passages, participants were asked to judge the similarity of the passages on a 5-point Likert scale and to explain the similarities and differences that guided their similarity rating. After the comparison task was completed, participants were asked to describe the principle that was central to the comparison passages. This multistage guided comparison-based schema abstraction task was intended to create a robust comparison condition that would fully engage the alignment process (see Kurtz et al., 2001).

After the study phase, the procedure in all conditions was exactly the same. All materials were removed and participants began the test phase without an intervening task or delay. The test phase was distributed on paper and presented the transfer problem with a space to provide a free response answer (see Appendix A for transfer problem). Care was taken by the experimenter to avoid any indication that the two tasks were related (recall that participants were completing a series of tasks during their experimental session). Participants were prompted to "Please read the following passage carefully and thoughtfully, and then proceed to the accompanying question." Upon completion, participants filled out a survey on their familiarity with the problem. Participants who indicated that they knew the solution to the problem and produced an answer that was scored as correct were excluded from the analysis ($n = 6$). Two additional participants were removed from the analysis because of experimenter error.

Responses were scored by two independent raters who were blind to condition. The raters created an independent score for each participant's response and (in the case of disagreement) the raters discussed the response and generated an agreed-upon score. The raters agreed on the scoring of 93% of the responses before discussion. Responses were scored as correct if the participant provided a solution that matched the principle presented in the learning phase. Specifically, a correct response was defined as recognizing that a virus can avoid the immune system of a host through the use of disguise, trickery, or an explicit attempt to change its appearance.

Results and Discussion

To address the question of how study task and example type influenced the rate of spontaneous transfer, a logistic regression model (R Core Team, 2015) was built with study task (category construction, comparison) and example type (solved, unsolved) included as categorical predictor variables. Results show that both study task and example type were reliable predictors of spontaneous transfer (see Figure 1). Participants receiving the category construction task were more likely ($M = .56$) than those receiving

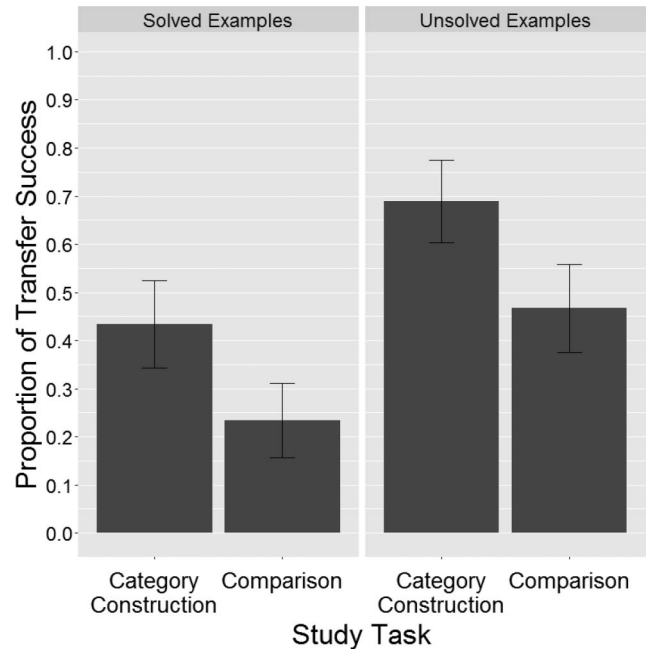


Figure 1. Proportion of responses adhering to the transfer principle produced during the spontaneous transfer test in Experiment 1a. The left frame displays the results from the Solved Examples group and the right frame displays the results from the Unsolved Examples group. Error bars depict 95% binomial confidence intervals (Dorai-Raj, 2014).

standard comparison ($M = .35$) to solve the test problem with the target principle (β Estimate = 0.92, $SE = 0.39$, Wald $Z = 2.362$, $p = .018$). This result validates our prediction that category construction is a powerful method for promoting spontaneous transfer. We also found that participants who received problems without solutions ($M = .58$) were reliably more likely than participants who received problems with solutions ($M = .33$) to show successful transfer (β Estimate = 1.06, $SE = 0.39$, Wald $Z = 2.705$, $p = .007$). The interaction between example type and study task was not a reliable predictor of transfer success.

It is possible that performance on the questions that were embedded in the study task might help interpret the pattern of results described above. We examined participants' accuracy on the category construction task (i.e., whether or not they produced the intended sort), similarity ratings in the comparison task, and explanations of the target principle as predictors of transfer. As it turned out, none of these factors were reliable predictors in logistic regression models built to predict spontaneous transfer frequency. From this we are led to believe that it is the act of engaging in the particular study tasks more than the quality of performance on the tasks that drives transfer outcomes. To elaborate, the data suggest that participants who failed to achieve the desired result of the study task (i.e., not producing the correct sort, not reporting high similarity between compared cases, or not being able to effectively justify their responses) did not benefit less from the study tasks as treatments against inert knowledge. This, in fact, suggests a certain robustness of the study tasks that is favorable to real-world application.

As a further analysis, it is worth considering whether the groups took different amounts of time and whether time on task was

predictive of transfer performance. We did not specifically control for time-on-task for two reasons: (a) we wanted to let the study experience of each participant occur naturally without an artificial cut-off and (b) despite decades of research we know of no evidence that spontaneous transfer can be improved by simply spending a bit more time with the study materials. The duration of the entire session (study and transfer phases) was recorded for each participant—note that this is a rough estimate of task time differences between conditions as it includes time spent working on the transfer problem. Because of the presence of missing task time data ($n = 8$), missing values were imputed with the *mi* package in R (Su, Gelman, Hill, & Yajima, 2011). The analysis shows that there was no effect of task completion time on transfer success ($p = .79$). The comparison group ($M = 14.44$ min) completed the entire session faster than category construction ($M = 18.17$ min; β Estimate = -3.673 , $SE = 1.419$, $t = -2.588$, $p = .012$). There was no difference in task time between the example type conditions ($p = .38$), so it did not take significantly longer to complete the task when the study cases were presented without solutions. While the category construction task takes longer to complete than comparison-based schema abstraction, transfer success is not predicted by time on task.

The present results show clear advantages for learning through category construction, as well as for the use of study problems without solutions. The combination of category construction and unsolved problems appeared to be especially effective—in fact close to the level of performance observed in previous work where explicit hints to use the examples from the study phase were provided (e.g., Catrambone & Holyoak, 1989; Gick & Holyoak, 1983; Spencer & Weisberg, 1986); however, there was not statistical support for an interaction. We know of no precedent for the present results in terms of: (a) the sheer rate of spontaneous transfer performance observed ($M = .69$ for category construction with unsolved cases); and (b) finding a reliable advantage for a novel study technique over comparison-based schema abstraction. It is worth noting that the best performing group (category construction with unsolved cases) received no statement of the solution principle at any point during the experiment, no confirmed identification of any particular case as a positive example of the target principle, and no confirmed identification (or implication) that any two cases share the same target principle. Ostensibly, such high transfer rates might be attributable to the use of a low-complexity and likely familiar principle (as well as the minimal context change for the transfer test), but the lower performance in the comparison conditions belies this interpretation.

We were surprised by the veritable gulf in transfer performance between the best (category construction with unsolved cases) and worst (the established best practice of comparing solved cases) study conditions—our inclination is to treat these results cautiously. The results clearly demonstrate that the use of category construction and the use of unsolved cases are promising ways to promote spontaneous transfer—and the significant improvement relative to the established standard of comparing solves cases is a major advance if it stands up to further evaluation. Because these findings hold considerable consequence toward theoretical and applied goals in the field, a high-powered replication is necessary to more definitively interpret the observed differences. This has been carried out (as Experiment 2, below), but first we report the

remaining results of our initial investigation into the type of contrast cases used in the category construction task.

Experiment 1b

As previously discussed, there are options for what type of contrast cases to use in a category construction task. In Experiment 1a, the category construction activity included three cases instantiating the target principle and three contrast cases that were related to the positive cases save for one key alignable difference (using disguise as a strategy for attack vs. using disguise as a strategy for defense). It is possible that category construction would be better served by maximizing the distinction between the analogs and the contrast cases. Although we have not seen evidence of the importance of sort success (relative to the importance of simply engaging in the sort), it could be helpful to have a higher rate of successful sorting. On the flip side, there could be a benefit of a desirable difficulty at work. It is also possible that there could be a benefit to having the target principle be the only coherent principle presented at study (as is typical in the A–B paradigm). Little is known about the impact of presenting examples of two coherent principles in the same study session—successful transfer of the target principle could be blocked by interference, competition, or confusion with a similar but competing principle.

Guided by these considerations, we developed a new set of contrast cases with the goal of minimizing surface or relational similarities to one another (low coherence) or to the positive cases of the target principle (see Appendix D). Our design was based on the hope of achieving the highest possible transfer rate, so we used the unsolved problem cases that showed such promise in Experiment 1a. In addition, the relatively low performance level for the comparison condition with the materials and procedures of Experiment 1a motivated us to include a baseline condition to more clearly verify the value of both comparison and category construction in promoting spontaneous transfer. The baseline condition was a sequential summarization task commonly used as an information-equivalent control alongside comparison-based schema abstraction tasks (Catrambone & Holyoak, 1989; Gentner, Loewenstein, Thompson, & Forbus, 2009; Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 1999).

Method

Participants and materials. Participants ($N = 121$) were drawn from the participant pool at Binghamton University and randomly assigned to condition. Each individual received partial fulfillment of a course requirement as compensation for participation. There were three between-subjects conditions: category construction, comparison-based schema abstraction, and sequential summarization. The materials were the same as in Experiment 1a except for the new set of contrast cases in the category construction condition. The two cases for sequential summarization were the same as those used in the comparison condition.

Procedure. Participants in the category construction study task ($n = 41$) were given a mat and six cards: three cards with examples that represented the target principle and three that were incoherent (see Appendix D). All examples were presented in the unsolved problem format. The instructions were modified slightly to orient participants toward constructing one meaningful category:

You are about to read a set of cards with passages that demonstrate different problem situations. Some of these passages are problems that can be solved using the same general principle or solution strategy. Your job is to pick out and put into a pile the problem situations that can be solved using the same general principle. Please consider the passages carefully to make sure you are satisfied with your sort.

After the task was completed, participants were prompted to explain the principle that was common to the cards in the “shared solution” group. Participants in the comparison study task ($n = 40$) received the same materials and instructions as in Experiment 1a. Participants in the summarization study task ($n = 40$) were given the same two positive cases of the target principle as in the comparison task, but on separate sheets of paper. They were prompted to read the passages carefully and thoughtfully and to read them well enough that they could restate the main points to a friend later on. Participants were asked to produce a written summarization after reading each example.

The test phase was identical to Experiment 1a. Participants who indicated that they were familiar with the transfer problem and its answer and provided the correct solution were excluded from analyses ($n = 5$). Additionally, one participant was removed from the analysis because of an apparent lack of English fluency. Responses were scored in the same manner as Experiment 1a. The raters’ initial scores were the same for 92% of the scored responses.

Results and Discussion

Successful transfer of the principle was used as the dependent variable in a logistic regression model with condition as an independent predictor variable. Results (see Figure 2) show that the category construction condition ($M = .46$; β Estimate = .952, $SE = 0.48$, Wald $Z = 1.979$, $p = .048$) and the comparison condition ($M = .48$; β Estimate = .999, $SE = 0.48$, Wald $Z = 2.066$, $p = .039$) reliably predicted transfer success as compared with the summarization baseline ($M = .25$). The category construction technique did not show an advantage over standard comparison with the different approach to the contrast set.

As in Experiment 1a, secondary analyses were conducted on study phase performance variables to further explain the spontaneous transfer results. A strong, but not overwhelming, majority of category construction participants (70%) created the intended grouping. This suggests that the use of incoherent contrast examples in category construction can increase the frequency of the intended example grouping. However, producing the accurate grouping once again did not appear to increase spontaneous transfer. Of those who sorted correctly, only 50% provided transfer solutions that corresponded to the principle (this was not a reliable predictor of transfer, $p = .49$). The category construction group ($M = .66$) was more likely than the comparison group ($M = .33$) to accurately describe the principle when cued (β Estimate = 1.39, $SE = 0.47$, Wald $Z = 2.942$, $p = .003$). This result is further evidence that a simple principle can be induced from multiple examples framed as problems with no inclusion of solution information. Considering that this difference was not found in Experiment 1a, we are led to conclude that incoherent contrast examples are helpful for cued production of a relational principle when learning with category construction. The remainder of the study phase questions did not produce reliable differences between study

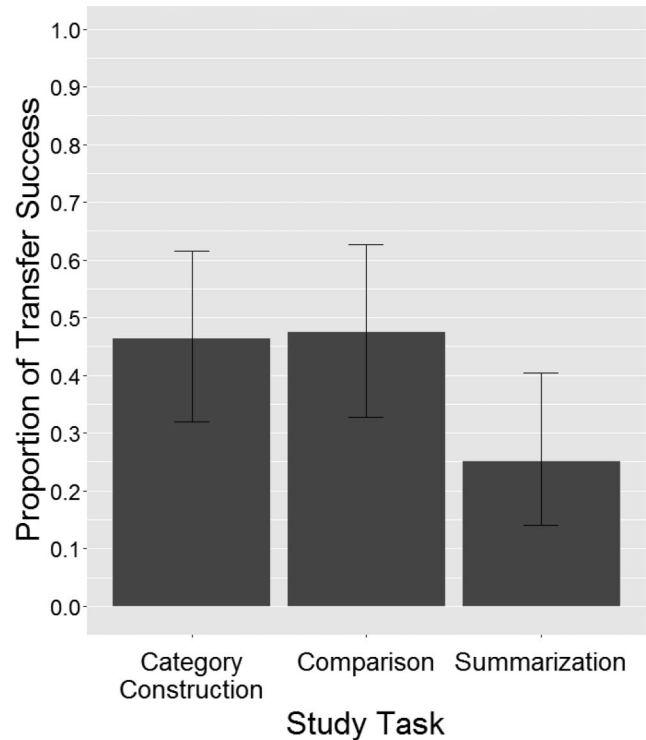


Figure 2. Proportion of responses adhering to the transfer principle produced during the spontaneous transfer test in Experiment 1b. Error bars depict 95% binomial confidence intervals (Dorai-Raj, 2014).

tasks: similarity ratings and quality of similarity and difference listing in the comparison condition, category construction accuracy, description of the relational principle, and familiarity with the transfer problem were not reliable predictors of transfer ($ps > .1$). Time-on-task (missing timing values imputed, $n = 5$) did not predict transfer success ($p = .32$) and the tasks did not differ in completion time for the entire session (category construction vs. comparison: $p = .12$; category construction vs. summarization: $p = .35$).

In our initial study (Experiment 1a), we saw distinctly low transfer after comparison of unsolved cases and distinctly high transfer after sorting unsolved cases. In the present experiment, we sought to evaluate the impact of a different type of contrast cases on category construction and introduced a summarization baseline group. The significant advantage for category construction relative to the summarization baseline supports the claim that category construction is an effective study task to promote transfer. However, the lack of reliable separation from the comparison group suggests that the use of incoherent contrast cases may not be the optimal way to set up the task if the ultimate goal is spontaneous transfer, especially in conjunction with the use of unsolved problems as learning examples. We speculate that the use of incoherent contrast cases in the form of problem situations without solutions may have somewhat undermined the goal of promoting a knowledge representation for the target principle with category status. Further research is needed to establish whether an incoherent contrast set can be more effective with solved cases or if it is simply best to stick with a coherent basis for the contrast cases.

The significant advantage for comparison relative to the summarization baseline supports and extends the use of comparison-based schema abstraction to promote transfer. Notably, this is the first demonstration (to our knowledge) of the advantage of traditional comparison over summarization with unsolved cases. Finally, the use of the baseline here helps to situate the results of Experiment 1a—confirming that the materials and procedures used in this series are appropriate to revealing facilitative effects of both comparison and category construction on spontaneous transfer.

Experiment 2

We now revisit the surprising findings that arose in Experiment 1a: (a) the unprecedented level of spontaneous transfer from the category construction group with unsolved cases, (b) the low level of transfer performance from the comparison group with solved cases; and (c) the significant main effects for category construction over the standard best practice of comparison-based schema abstraction and unsolved cases over solved cases. Our experience using the A–B paradigm with its binomial outcome measure (transfer success) suggests that repeated testing using traditional sample sizes and the exact same treatment tends to show considerable variability—and, further, that the reliability of group differences can hang on the performance of just a few participants. Additionally, a sample size increase offers greater potential to better reveal how performance during the study phase predicts transfer (the secondary measures that have not been shown to be reliable above). The goal of the present experiment was to use a minimal design with a larger sample size to produce a clear and definitive outcome for the following question: Does category construction with unsolved cases and a coherent contrast set lead to better transfer than comparison-based schema abstraction? Note that we are not committed to linking category construction and unsolved cases, but an important goal in this work is to figure out how to generate the highest possible transfer rate and the conjunction of the two showed the greatest promise in Experiment 1a. It is likely that unsolved problems will prove to be unsuitable for more complex principles and different levels of learners—this observation has borne out in the results of a parallel investigation of the effectiveness of category construction in 7th grade science classrooms where added supports were critical for sort success (Honke et al., 2015).

Our design included two comparison groups (solved vs. unsolved cases) to address the unresolved questions arising from the low performance of the comparison with solved examples group in Experiment 1a and to provide further clarity with regard to the effectiveness of these alternatives for comparison-based schema abstraction. We did not include a category construction with solved examples group so as to meet our goal of a reduced design with maximum power and because the most dramatic findings from Experiment 1a did not hinge on this condition. Following Simonsohn's (2015) proposal for designing replication studies, we targeted a sample size 2.5 times greater than that of Experiment 1a.

Method

Participants and materials. Participants ($N = 292$) were Binghamton University students recruited from the participant

pool and compensated with partial fulfillment of a course requirement. Participants were randomly assigned to one of three between-subjects conditions: category construction with unsolved examples ($n = 94$), comparison with unsolved examples ($n = 97$), comparison with solved examples ($n = 101$). The learning examples and transfer problem were identical to those in Experiment 1a (see Appendices A and C).

Procedure. The procedure was just as in Experiment 1a with only the slight deviation that an additional YES/NO question was added to the posttask familiarity survey. The additional item asked participants whether or not they recognized any similarity between the passages in the study phase of the experiment and the transfer test passage. As in the experiments above, participants who indicated that they knew the solution to the problem and provided an answer scored as correct were excluded from the analysis ($n = 5$). Three additional participants were removed from the analysis because of experimenter error. Scoring was conducted in the same manner; although this time we used raters who were naïve to the experiment hypotheses in addition to being blind to condition. Before discussion, the raters agreed on 93% of the scoring assignments.

Results and Discussion

A logistic regression model with study task included as a categorical predictor was built to analyze the effect of category construction and comparison on transfer. The transfer advantage observed in Experiment 1a for category construction over comparison was replicated (see Figure 3). The analysis uncovered a category construction advantage where category construction ($M = .44$) was a reliable predictor of transfer success when compared with comparison with unsolved examples ($M = .27$; β Estimate = $.7479$, $SE = 0.31$, Wald $Z = 2.42$, $p = .016$), and comparison with solved examples ($M = .29$; β Estimate = 0.6527 , $SE = 0.30$, Wald $Z = 2.16$, $p = .031$). It is important to note that we no longer observe a skyrocketing transfer rate for category construction with unsolved examples. However, we do see a compelling replication of the major finding: a significant advantage for category construction over traditional comparison. In addition, we do not replicate the advantage for unsolved cases over solved cases with traditional comparison. It appears that the “low” transfer performance for comparison with solved cases in Experiment 1a was actually right on target—the results with the larger sample suggest that the previous performance of the other groups on this task were inflated (see Table 2).

Turning now to the secondary analysis of study phase data, recall that the related analyses of Experiments 1a and 1b suggested that performance on the study phase did not predict spontaneous transfer. In Experiment 2, however, producing the intended categories was marginally significant as a predictor of transfer (β Estimate = 0.739 , $SE = 0.43$, Wald $Z = 1.736$, $p = .083$), but only 53% of the category construction participants produced this outcome (an expected drop from the rate of correct sorts with incoherent contrast examples in Experiment 1b). From these results it is less clear that performance in the task is not associated with transfer outcomes. As such, it is possible that a greater focus on supports for successful sorting in the task could yield better transfer outcomes, but these supports must be carefully designed to retain aspects of the task that promote independent discovery of

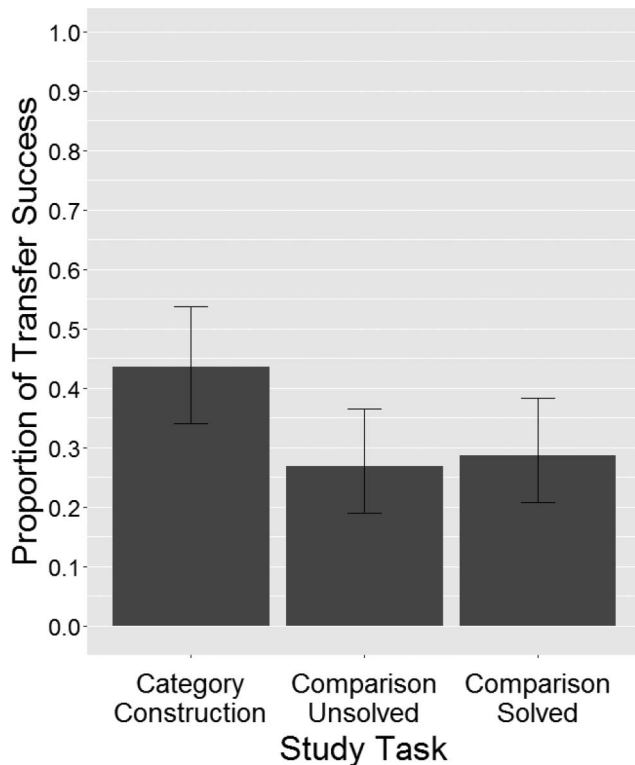


Figure 3. Proportion of responses adhering to the transfer principle produced during the spontaneous transfer test in Experiment 2. Error bars depict 95% binomial confidence intervals (Dorai-Raj, 2014).

the relational similarities and underlying principle of the learning examples.

Total task completion time was analyzed to address the possibility that the effects of category construction might be attributable to a longer study period. Task time was not a reliable predictor of transfer success across conditions ($p = .92$). There were overall

completion time differences—category construction took longer to complete ($M = 18.64$ min) than comparison with solved examples ($M = 14.47$ min; β Estimate = 4.22, $SE = 0.72$, $t = 5.879$, $p < .001$) and comparison with unsolved examples ($M = 14.52$ min; β Estimate = 4.164, $SE = 0.73$, $t = 5.741$, $p < .001$). In summary, category construction took longer than either comparison task, but task time did not predict transfer (even when accounting for the variance predicted by condition). Given that there is no a priori basis to believe that an increase in study time promotes spontaneous analogical transfer and the present evidence shows that time on task was not a significant predictor, the possibility that the category construction advantage is because of increased time appears remote.

Next, we examine participants' cued descriptions of the relational principle. Accurate descriptions of the principle during the learning phase reliably predicted transfer (β Estimate = 0.816, $SE = 0.27$, Wald $Z = 3.040$, $p = .002$). Looking between learning conditions, the comparison group that received unsolved examples produced fewer accurate principle descriptions ($M = .32$) than the category construction group ($M = .70$; β Estimate = 1.611, $SE = 0.31$, Wald $Z = 5.142$, $p < .001$) or the comparison group that received solved examples ($M = .77$; β Estimate = 2.025, $SE = 0.32$, Wald $Z = 6.262$, $p < .001$). The latter two groups were not reliably different. This is compelling evidence that category construction participants produced relationally similar solutions to the sorted problems even though that was not the goal of the task. The comparison group that did not receive the solutions was worse at deriving the solution, similar to the results of Experiment 1b.

Finally, participants who answered "Yes" to the question asking if they noticed any similarities between the study and test phases were more likely to provide the disguise strategy as an answer to the transfer problem (β Estimate = 1.289, $SE = 0.32$, Wald $Z = 3.975$, $p < .001$); the category construction group was more likely to answer Yes to this question than the (collapsed) comparison groups (β Estimate = 0.683, $SE = 0.29$, Wald $Z = 2.331$, $p = .02$). We take this as further evidence that category construction is more effective than comparison-based schema abstraction in pro-

Table 2
Descriptive Statistics and Sample Size by Condition

Condition	Example type	N	Proportion successful transfer (count)	95% Confidence interval	
				Lower	Upper
Experiment 1A					
Category construction	Solved	$n = 30$.69 (20)	.484	.810
	Unsolved	$n = 29$.43 (13)	.281	.628
Comparison	Solved	$n = 30$.47 (14)	.299	.641
	Unsolved	$n = 30$.23 (7)	.116	.415
Experiment 1B					
Category construction	Unsolved	$n = 41$.46 (19)	.319	.615
Comparison	Unsolved	$n = 40$.48 (19)	.327	.627
Summarization	Unsolved	$n = 40$.25 (10)	.140	.405
Experiment 2					
Category construction	Unsolved	$n = 94$.44 (41)	.339	.538
Comparison	Unsolved	$n = 97$.27 (26)	.189	.365
	Solved	$n = 101$.29 (29)	.207	.383
Experiment 3					
Category construction	Unsolved	$n = 112$.51 (57)	.417	.600
Reading comprehension with retrieval practice	Unsolved	$n = 112$.41 (46)	.324	.504

moting the recognition of analogous problems between study and test.

Experiment 3

The results of the previous experiments support the category status hypothesis—that is, that experiencing examples of a relational principle in terms of category membership promotes abstraction and increases spontaneous transfer; however, there are alternative explanations to consider. We have not addressed the possibility that the category construction task offers some type of advantage that is unrelated to developing a more category-like knowledge representation. For example, the task could be effective because of the availability of more and/or different types of (un-guided) comparison opportunities because it includes an additional positive example and three counterexamples that are not present in traditional comparison-based study. Further, the additional examples could provide a basis for successful reminding that is entirely case-based (despite the intended lack of surface similarity among the set of cases included in the experiment) and not driven by the formation of an abstract representation at all. To address these concerns, we tested category construction against a reading comprehension task that included all of the content available in the category construction task but lacked the sort activity (consisting of actively linking examples to common or distinct categories). Therefore, each condition includes exposure to the same number of cases exemplifying the target principle as well as the same number of counterexamples of a contrasting principle.

It is important to be clear about the strengths and weaknesses (the latter being unavoidable as far as we can tell) of this approach. To begin, we should address why controlling for matching content between the conditions was not part of the original experiments (Experiment 1 and Experiment 2). Under the category status hypothesis, we expect an advantage because of experiencing a third positive example (this offers a stronger basis for forming a category) and because of experiencing counterexamples (helping to specify the category boundary). Therefore, a control group matched for content with the category construction task receives a very significant push toward category status in its own right—and, therefore, serves only as a test of a moderate versus strong push toward increased category status. It may be quite unlikely for our available dependent measures (and sample sizes) to provide sufficient sensitivity to detect a reliable difference between moderate and strong levels of category-like representation. In point of fact, the comparison condition used in Experiment 1 and Experiment 2 also invites category status to a degree (via multiple positive examples plus the guided comparison task) but at a substantially weaker level. Accordingly, we expected and observed an effect in those experiments, but the present design may provide too fine of a discrimination to reveal an effect. In summary, a weakness of the present design toward assessing the category status hypothesis is that the most likely outcome under the hypothesis is a small numerical advantage for the category construction group—this could create the appearance of a failure of the hypothesis when that is hardly the case.

As it turns out, the problem is further exacerbated. While the reading comprehension control is the best option we were able to identify, this study task provides subjects with a set of advantages that substantially bias the experiment against the category con-

struction group. First, a clear indication was made to control subjects that the study phase was relevant to subsequent testing (contrary to a truly spontaneous test) to ensure an adequate level of engagement and confirm that the study materials were encoded. In addition, this required an assessment at the end of the study phase that amounted to a series of retrieval practice trials (for evidence on the efficacy of this procedure; see Roediger & Karpicke, 2006). Furthermore, the sequential presentation of items suggested particular pairwise comparisons (as opposed to being maximally unguided in category construction) and we deemed it necessary to give explicit instructions to consider the potential solution for each problem (as opposed to not doing so in the category construction condition). Lastly, to maintain maximal alignment between the conditions, we had to make a modification of unknown impact on the category construction task by including distractor items in the sort set. The extent to which each of these factors favors the control condition is unknown, but these are by no means trivial considerations. To reiterate, we are aware of no experimental design that achieves the present experimental goals without these complicating factors.

In light of this panoply of factors, this experiment may in fact be of limited worth as a test of whether or not there is a category construction advantage. However, we put forth that the experiment is a valuable addition to the series for a number of reasons. First, a significant effect could emerge. This would suggest that even a nuanced increase in the degree of inviting category status is powerful enough to impact outcomes—even on a tilted playing field. Second, the design offers guidance in interpreting the results of the previous experiments. A specific concern is that the category construction group may have benefitted from studying more cases for a reason other than contributing to category status. If category-level representation is not a factor, then the likelihood of transfer depends on similarity-based retrieval of cases. On this view, encoding more cases provides another opportunity for successful access and elicits a somewhat broader range of similarity-based matching potential. If this is the mechanism at work, it follows that memory for the cases would be a strong predictor of transfer success. The reading comprehension control allows us to assess whether or not a purely case-based account offers a coherent interpretation of the transfer phenomenon.

Method

Participants and materials. Participants ($N = 227$) were Binghamton University students recruited from the participant pool and compensated with partial fulfillment of a course requirement. As in the previous experiments, participants who indicated that they knew the solution and provided the solution from the study phase were excluded from the analysis ($n = 3$). Participants were randomly assigned to category construction ($n = 112$) or reading comprehension ($n = 112$). The learning examples and transfer problem were identical to those in Experiment 1a; two distractor examples from Experiment 1b were also included in the study phase (see Appendix E). Sample size was based on previously observed effect size and simulation-based power analysis, though few reference points existed as a basis for effect size predictions in this context. The reading comprehension control condition provided a number of elements hypothesized to increase category status: experience of more examples and counterexam-

ples, unsolved problems as study material, temporal juxtaposition in-line with past work on comparison (e.g., Blanchette & Dunbar, 2002). While these features match the conditions of the category construction activity, the reading comprehension activity also provided an explicit instruction to study for test, retrieval practice, and structured (temporal) between-category contrast of learning examples (Rittle-Johnson & Star, 2007, 2009; Schwartz & Bransford, 1998). We maintain the hypothesis that the active process of category construction is superior to these alternative routes to category status but the critical question is how much. A 10-point condition advantage (β Estimate = 0.40, $1 - \beta = .80$, $p < .05$) reaches the floor of acceptable power levels at $N = 220$ (e.g., 80%; see Lakens, 2014; analysis with WebPower, Zhang & Yuan, 2018). Shrinking this advantage to 5% more frequent transfer—a practically important difference for the issues at study—dramatically increases suggested sample size to $N = 800$ (β Estimate = 0.20, $1 - \beta = .80$, $p < .05$). We recruited participants based on the design parameters suggested by the first analysis, 220 participants with a small addition to address experimenter error and participants' a priori problem knowledge.

Procedure. The experiment closely matched the procedures presented above. Participants were seated at private desks and experiment materials were provided on paper. The category construction task only differed from the previous experiments in that two unrelated examples were included as to-be-categorized examples. Therefore, the goal of the Experiment 3 category construction task was to identify and remove these unrelated distractors while also forming two groups solvable by the same general principle. The reading comprehension condition was provided with the same eight examples (three principle-consistent passages, three principle-inconsistent passages, and two unrelated passages) as the category construction task, but the goal was to read each passage, think about possible solutions and prepare to be tested on the content of the passages. The final page of the reading comprehension study task presented four free response questions that targeted superficial aspects of the study passages (participants were not allowed to refer back to the examples during the reading comprehension assessment). After the study phase, participants were presented with the transfer problem. No indication was given that the study and transfer phases of the experiment were related. Finally, a familiarity survey was administered that was identical to that of Experiment 2 with one important addition—participants were asked to recall a specific principle-consistent example from the study phase and reproduce the passage to the best of their ability. Responses to the transfer problem, the sort explanation (category construction condition only), the reading comprehension questions (reading comprehension condition only) and the study phase memory assessment were scored by raters naïve to condition (where possible) and the experiment's hypothesis. Before discussion, the raters agreed on 90% of the spontaneous transfer score assignments.

Results and Discussion

We tested for the possibility of a transfer advantage conferred by the category construction task relative to a closely matched reading comprehension task. A GLM built with study phase as a categorical predictor uncovered no reliable difference in spontaneous transfer frequency between the category construction ($M =$

.51) and reading comprehension ($M = .41$) conditions, β Estimate = 0.40, $SE = 0.27$, Wald $Z = 1.472$, $p = .14$).

Consistent with our assertion that memory for concrete instantiations of the relational principle is not the key driver of spontaneous transfer (cf. Ross, 1984, 1987), recall accuracy for study phase materials did not predict transfer ($p = .18$). Accuracy scores on the cued recall task were found to differ between conditions—the category construction group had *lower* accuracy in recalling details of study passages as compared with the reading comprehension group (beta Estimate = 0.40, $SE = 0.19$, $t = 2.08$, $p = .039$). This evidence suggests that successful spontaneous transfer cannot be attributed to better memory for concrete examples.

As above, accurate performance on the tasks in the study phase can be analyzed as predictors of transfer success. For the category construction group, accurate sorting was a reliable predictor of spontaneous transfer (beta Estimate = 1.23, $SE = 0.46$, Wald $Z = 2.694$, $p = .007$). However, sort accuracy was considerably lower than previous experiments ($M = .29$)—perhaps as a result of the added difficulty of considering and excluding unrelated passages. Providing an accurate description of the relational principle shared by the target examples also predicted transfer, though this effect was only marginally significant (beta Estimate = 1.24, $SE = 0.64$, Wald $Z = 1.953$, $p = .051$). These results suggest that the previously elusive effect of accurate performance on the category construction task—particularly sort accuracy—is important for spontaneous transfer. Producing an accurate sort and accurately describing the underlying relational principle for the relationally similar examples independently predict the likelihood of spontaneous transfer.

For the reading comprehension group, a majority of participants (70%) had perfect accuracy on the reading comprehension assessment questions. Accurate recall on the assessment did not predict transfer success ($p = .77$); this is further evidence that the nature of the encoded representation—relational category versus encoding via serial presentation of relationally similar passages—is critical for spontaneous transfer.

Overall, accurate recall immediately after study (reading comprehension group only) and after the spontaneous transfer test (both groups) did not reliably predict spontaneous transfer. This outcome makes it difficult to attribute the success of the category construction task to better memory for the content presented in the study phase. This outcome fits quite well, however, with the predictions of the category status hypothesis—it suggests that encoding the examples as members of a relational category helps to produce an abstract category representation that is more likely to be activated when a novel problem is experienced.

General Discussion

The primary goal of these studies was to introduce and test the category construction technique for promoting spontaneous transfer. The approach presented here is the result of a new theoretical viewpoint—the category status hypothesis—based on the prediction that the problem of inert knowledge can be minimized by learning to view relationally similar examples as members of a category, that is, organizing seemingly distinct examples as a conceptual kind with a coherent basis and membership. The proposal that learning to think of cases as members of a category will promote portability, that is, access in situations that share high structural similarity and low

superficial similarity, is motivated by the fact that categorization is a ubiquitous, active, fluent, directed, and impactful cognitive process that drives our ability to make sense of what we experience. To address this hypothesis, we investigated the effect of guiding learners to discover and form categories of relationally similar examples through a sorting activity and set this technique against a comparison-based schema abstraction task and an information-matched reading comprehension task.

The results provide the first evidence that a learning technique designed to promote category status can have a profound impact on spontaneous transfer success rates. Experiment 1a uncovered an unprecedented rate of spontaneous transfer success for this domain (verbal problem solving) in the category construction condition. While we do not believe cross-study comparisons of successful transfer rates can be made without caveat, it is worth noting that the high level of purely spontaneous transfer success (i.e., uncued, no hints) observed in Experiment 1 is quite rare relative to the prior literature. The rate of spontaneous transfer observed in the category construction group approached that of cued transfer in prior research, for example, 68% success after comparison-based schema abstraction and a cue to use previous information (Experiment 1; Catrambone & Holyoak, 1989).

In Experiment 1b, using an incoherent contrast set, we found an advantage for category construction and comparison in relation to explanation through summarization alone, but category construction was not reliably better than standard comparison. Consideration of Experiments 1a and 1b together suggests that the optimal design of the category construction technique should include alignable contrast examples that form a coherent category (however, this could change with the use of solved cases). In Experiment 2, we conducted a replication of the key results of Experiment 1a with higher a priori power. We again found a higher rate of spontaneous transfer for the category construction group, though these results did not match the high transfer rate initially observed. There was no advantage for unsolved examples with comparison-based schema abstraction. In Experiment 3, we tested a modified category construction task (including distractors) against a reading comprehension task featuring retrieval practice and explicit direction to study for test. No advantage for spontaneous transfer was found for category construction, but the overall results suggest that the effectiveness of category construction cannot be explained by better memory for concrete examples. We also found more evidence that study task accuracy (sort accuracy, principle explanation) is important for spontaneous transfer success.

Taken together, these results send a clear message that the category construction approach is effective and, while further study is certainly needed, the evidence suggests that: (a) the sort task may be the best available approach to promoting spontaneous transfer of a mastered principle; and (b) unless there is a hidden benefit inherent in the sort task, then category status is an important factor in promoting transfer outcomes. Along these lines, it is important to consider potential alternate explanations of the success of the category construction condition. Such interpretations might attribute the benefits observed to one of the component elements of the sort task. For example, the simplest difference is that the sort task provides three examples of the target principle, as opposed to two in the comparison for schema abstraction condition. Providing more examples potentially widens the range of similarity-based matches. However, this is not a good explanation

of the present results because the match between our transfer problem and the provided sources is purely a relational match. Therefore, the content available for a relational match does not grow any stronger with the addition of a third case and there is no supporting superficial match between any of the sources and the transfer problem. Additionally, applied work addressing content in the area of physiology failed to reveal differences in cued transfer between study tasks that provide two and three analogous situations (Kulasegaram et al., 2017). Finally, we found that successful recall of source material did not predict transfer and the experience of three principle-consistent source passages in conjunction with the expectation of future testing was not enough to produce improved outcomes on the cued transfer test. We do believe that increasing the number of cases matters to transfer—through a mechanism of promoting category status—and suggest that the present findings of higher transfer rates are best understood as a combined effect of promoting category status via the sort task along with promoting category status by providing an additional positive example.

Another possible explanation could be that the six items in the sort task allow for more and different comparisons than the guided comparison of two analogs. In other words, one could see the sort task as a “souped-up” comparison opportunity rather than having anything to do with category status. This is an interesting interpretation because of the recognized power of comparison-based learning—there are three possible pairwise comparisons among analogs representing the target principle plus the potential benefits of learning from contrasting cases (Rittle-Johnson & Star, 2007, 2009; Schwartz & Bransford, 1998). However, the benefits of comparison are firmly understood to depend on an active, guided process rather than mere juxtaposition (Catrambone & Holyoak, 1989; Gentner et al., 2003; Kurtz et al., 2001) and robust comparison rarely occurs without explicit task supports (that are not present in the sort task). It would take considerably more time to complete the work of as many as 15 possible comparisons (time-on-task data suggests that category construction takes longer, but the added time—approximately four additional minutes for the entire session—is not consistent with the completion of multiple full comparisons). The sort task in fact has very little to offer when viewed from the perspective of the memory-based account: instead of providing a clear, directed comparison of two positive cases, the learner has no guided comparison task—and no information regarding which items are positive cases of the same principle. Lastly, there is little reason to believe that multiple comparisons or contrasts would be beneficial to produce a clean abstract schema given the low-complexity principle featured in this research. Our interpretation is that the success of category construction in promoting transfer is not well accommodated by the prevailing view of schema retrieval but does accord well with the predictions entailed by the category status hypothesis.

As to the possible benefit of using unsolved problem examples during study, our results are less clear. Studying problems was reliably better than studying solutions in Experiment 1a. The highest rate of spontaneous transfer with standard comparison in our series of studies was observed in an unsolved problems condition (Experiment 1b). Ultimately, the results of Experiment 2 take precedence because of the larger sample—where the effect of including a solution with the study examples was not reliably different from excluding the solution. Considering related research

in this domain (Lockhart et al., 1988; Needham & Begg, 1991), perhaps the lack of an unsolved problem advantage can be attributed to the additional task demands in the present studies. In addition to receiving unsolved examples, participants had to complete the comparison or category construction activities—a requirement that was not present in previous research on the effect of using unsolved problems. This is consistent with recent evidence that competing task requirements appear to attenuate the learning advantages conferred by the comparison process (Patterson & Kurtz, 2015).

Finally, this work has outlined an important experimental method and design issue. Spontaneous transfer research using the A–B paradigm (and its dichotomous outcome variable) will need significant adjustment as practices change in response to contemporary views on the limitations of the Neyman-Pearson null-hypothesis significance testing paradigm. It was not uncommon in recent history to see experiments with fewer than 30 participants per condition. The severity of this issue is laid bare in Experiment 3, where a priori power analysis suggested a sample of 220 participants would be sufficient for a statistically reliable 10-point difference with 80% power. A smaller effect (5-point transfer advantage) would call for prohibitive resource expenditure in the context of an in-person laboratory environment.

Toward the wider context of this research, it is important to consider the broad educational implications of the inert knowledge problem and its effect on learners. Stated plainly, if this problem remains unresolved the educational system will continue to produce poor learning outcomes. The most recent National Assessment of Educational Progress Report (NAEP; National Center for Education Statistics, 2012) concretely outlines this problem, showing that only 32% of 8th grade students are at or above the proficient level in science. A major component of science proficiency is the ability to see when and how explanatory principles are pertinent in specific situations—exactly the type of knowledge that is targeted by the category construction technique. One of the notable successes in the science of learning is the discovery that comparing two analogs leads to significantly better spontaneous transfer than studying a single case or summarizing two separate cases or seeing a statement of the abstract principle (Ross & Kilbane, 1997). However, the scope of this improvement is not overwhelming on a practical scale. Now we have a technique that yields a significant advantage over the technique that yields a significant advantage—suggesting a qualitative leap in our ability to combat the problem of inert knowledge. One caveat is that our evidence is restricted to the case of transferring a straightforward (i.e., fully mastered) principle—that is the core concern of inert knowledge—but many kids in today's classrooms are not fully mastering the principles in their curriculum. Future work will determine whether the benefits of category construction extend to target principles that lie closer to the frontier of a learner's understanding.

Given the bleak outlook found in the NAEP report, it seems clear that the improvements observed because of category construction in these studies would surely be welcome if they translated to authentic education settings. Fortunately, this work has already begun. J. I. Smith et al. (2013) have found success using category construction as a means of assessing conceptual expertise in biology at the college level. We have begun testing category construction as an instructional technique for target

principles from the curriculum in 7th grade science classrooms. This has led to promising results for transfer of knowledge relative to a worksheet control group (Honke et al., 2015; Premo, Cavagnetto, Honke, & Kurtz, 2019). The best results were obtained with a sort task following up a lesson and using labeled sort groups with a Head Start of one sorted example. These projects provide a glimpse into the ways in which category construction can be used to make measurable gains in formal education.

In summary, this project addresses the impact of different study techniques on the rate of successful spontaneous transfer in a problem solving task. We set out to test the hypothesis that a learning method based on the novel theoretical framework of the category status hypothesis could most effectively overcome the problem of inert knowledge. The findings represent a compelling fulfillment of that prediction as the category construction technique does indeed promote transfer outcomes to a greater degree than the established best practice of comparison-based schema abstraction. Future work will explore the effectiveness of this technique and its underlying basis in the laboratory and authentic education settings.

References

- Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*, 87–113. <http://dx.doi.org/10.1080/00461520.2013.775712>
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128*, 612–637. <http://dx.doi.org/10.1037/0033-2909.128.4.612>
- Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166. <http://dx.doi.org/10.1037/0278-7393.15.1.153>
- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–206). Cambridge, MA: The MIT Press.
- Blanchette, I., & Dunbar, K. (2002). Representational change and analogy: How analogical inferences alter target representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 672–685. <http://dx.doi.org/10.1037/0278-7393.28.4.672>
- Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York, NY: Wiley.
- Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1020–1031. <http://dx.doi.org/10.1037/0278-7393.22.4.1020>
- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*, 355–376. <http://dx.doi.org/10.1037/0096-3445.127.4.355>
- Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1147–1156. <http://dx.doi.org/10.1037/0278-7393.15.6.1147>
- Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152. http://dx.doi.org/10.1207/s15516709cog0502_2
- Corral, D., & Jones, M. (2014). The effects of relational structure on analogical learning. *Cognition, 132*, 280–300. <http://dx.doi.org/10.1016/j.cognition.2014.04.007>

- Corral, D., Kurtz, K. J., & Jones, M. (2018). Learning relational concepts from within-versus between-category comparisons. *Journal of Experimental Psychology: General*, *147*, 1571.
- Day, S. B., & Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 551–567. <http://dx.doi.org/10.1037/a0022333>
- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist*, *47*, 153–176. <http://dx.doi.org/10.1080/00461520.2012.696438>
- Dorai-Raj, S. (2014). *binom: Binomial confidence intervals for several parameterizations* (R package version 1.1–1) [Computer software manual]. Retrieved from <https://CRAN.R-project.org/package=binom>
- Engle, R. A. (2012). The resurgence of research into transfer: An introduction to the final articles of the transfer strand. *Journal of the Learning Sciences*, *21*, 347–352. <http://dx.doi.org/10.1080/10508406.2012.707994>
- Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 141–205. http://dx.doi.org/10.1207/s15516709cog1902_1
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170. http://dx.doi.org/10.1207/s15516709cog0702_3
- Gentner, D. (1989). The mechanisms of analogical learning in similarity and analogical reasoning. In S. Vosnidou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 199–241). London: Cambridge University Press.
- Gentner, D. (2005). The development of relational category knowledge. In L. Gershkoff-Stowe & D. H. Rakison (Eds.), *Building object categories in developmental time* (pp. 245–275). Hillsdale, NJ: Erlbaum.
- Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*, 752–775. <http://dx.doi.org/10.1111/j.1551-6709.2010.01114.x>
- Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. Wolff (Eds.), *Categorization inside and outside of the lab: Festschrift in honor of Douglas L. Medin* (pp. 151–175). Washington, DC: American Psychological Association. <http://dx.doi.org/10.1037/11156-009>
- Gentner, D., Levine, S. C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., & Honke, G. (2016). Rapid learning in a children's museum via analogical comparison. *Cognitive Science*, *40*, 224–240. <http://dx.doi.org/10.1111/cogs.12248>
- Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*, 393–408. <http://dx.doi.org/10.1037/0022-0663.95.2.393>
- Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*, 1343–1382. <http://dx.doi.org/10.1111/j.1551-6709.2009.01070.x>
- Gentner, D., & Markman, A. B. (1995). Similarity is like analogy: Structural alignment in comparison. In C. Cacciari (Ed.), *Similarity in language, thought, and perception* (pp. 111–147). Brussels, Belgium: BREPOLs.
- Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology*, *25*, 524–575. <http://dx.doi.org/10.1006/cogp.1993.1013>
- Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*, 306–355. [http://dx.doi.org/10.1016/0010-0285\(80\)90013-4](http://dx.doi.org/10.1016/0010-0285(80)90013-4)
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. [http://dx.doi.org/10.1016/0010-0285\(83\)90002-6](http://dx.doi.org/10.1016/0010-0285(83)90002-6)
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125–157. [http://dx.doi.org/10.1016/0010-0277\(94\)90065-5](http://dx.doi.org/10.1016/0010-0277(94)90065-5)
- Goldstone, R. L., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466. [http://dx.doi.org/10.1016/S0010-0285\(02\)00519-4](http://dx.doi.org/10.1016/S0010-0285(02)00519-4)
- Goldstone, R. L., & Son, J. Y. (2005). The transfer of scientific principles using concrete and idealized simulations. *Journal of the Learning Sciences*, *14*, 69–110. http://dx.doi.org/10.1207/s15327809jls1401_4
- Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition*, *137*, 137–153. <http://dx.doi.org/10.1016/j.cognition.2014.12.001>
- Goldwater, M. B., Markman, A. B., & Stilwell, C. H. (2011). The empirical case for role-governed categories. *Cognition*, *118*, 359–376. <http://dx.doi.org/10.1016/j.cognition.2010.10.009>
- Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin*, *142*, 729.
- Gonzalez, C., & Wong, H. (2012). Understanding stocks and flows through analogy. *System Dynamics Review*, *28*, 3–27. <http://dx.doi.org/10.1002/sdr.470>
- Greeno, J. G., Smith, D. R., & Moore, J. L. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and instruction* (pp. 99–167). Norwood, NJ: Ablex.
- Higgins, E. J. (2017). The complexities of learning categories through comparisons. In B. H. Ross (Ed.), *The psychology of learning and motivation: The psychology of learning and motivation* (Vol. 66, pp. 43–77). San Diego, CA: Elsevier Academic Press. Retrieved from <https://psycnet.apa.org/record/2017-28922-002>
- Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition*, *15*, 332–340. <http://dx.doi.org/10.3758/BF03197035>
- Honke, G., Cavagnetto, A. R., Kurtz, K. J., Patterson, J. D., Conaway, N. B., Tao, Y., & Marr, J. C. (2015). *Promoting transfer and mastery of evolution concepts with category construction*. Paper presented at the American Educational Research Association annual meeting, Chicago, IL.
- Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, *69*, 596–608. <http://dx.doi.org/10.1037/h0021980>
- Imai, S., & Garner, W. R. (1968). Structure in perceptual classification. *Psychonomic Monograph Supplements*, *2*(Whole No. 25).
- Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological Bulletin*, *112*, 500–526. <http://dx.doi.org/10.1037/0033-2909.112.3.500>
- Kulasegaram, K. M., Chaudhary, Z., Woods, N., Dore, K., Neville, A., & Norman, G. (2017). Contexts, concepts and cognition: Principles for the transfer of basic science knowledge. *Medical Education*, *51*, 184–195. <http://dx.doi.org/10.1111/medu.13145>
- Kurtz, K. J. (2015). Human category learning: Toward a broader explanatory account. *Psychology of Learning and Motivation*, *63*, 77–114. <http://dx.doi.org/10.1016/bs.plm.2015.03.001>
- Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1303–1310. <http://dx.doi.org/10.1037/a0031847>
- Kurtz, K. J., & Gentner, D. (2013). Detecting anomalous features in complex stimuli: The role of structured comparison. *Journal of Experimental Psychology: Applied*, *19*, 219–232. <http://dx.doi.org/10.1037/a0034395>
- Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences*, *10*, 417–446. http://dx.doi.org/10.1207/S15327809JLS1004new_2
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, *44*, 701–710. <http://dx.doi.org/10.1002/ejsp.2023>
- Lave, J. (1988). *Cognition in practice: Mind, mathematics and culture in everyday life*. Cambridge, England: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511609268>

- Lobato, J. (2012). The actor-oriented transfer perspective and its contributions to educational research and practice. *Educational Psychologist*, 47, 232–247. <http://dx.doi.org/10.1080/00461520.2012.693353>
- Lockhart, R. S., Lamon, M., & Gick, M. L. (1988). Conceptual transfer in simple insight problems. *Memory & Cognition*, 16, 36–44. <http://dx.doi.org/10.3758/BF03197743>
- Loewenstein, J. (2010). How one's hook is baited matters for catching an analogy. *Psychology of Learning and Motivation*, 53, 149–182. [http://dx.doi.org/10.1016/S0079-7421\(10\)53004-4](http://dx.doi.org/10.1016/S0079-7421(10)53004-4)
- Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review*, 6, 586–597. <http://dx.doi.org/10.3758/BF03212967>
- Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition*, 24, 235–249. <http://dx.doi.org/10.3758/BF03200884>
- Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Artificial Intelligence*, 13, 329–358.
- Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology*, 19, 242–279. [http://dx.doi.org/10.1016/0010-0285\(87\)90012-0](http://dx.doi.org/10.1016/0010-0285(87)90012-0)
- Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 407–415. <http://dx.doi.org/10.1037/0278-7393.30.2.407>
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533. [http://dx.doi.org/10.1016/S0022-5371\(77\)80016-9](http://dx.doi.org/10.1016/S0022-5371(77)80016-9)
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: The MIT Press. <http://dx.doi.org/10.7551/mitpress/1602.001.0001>
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- National Center for Education Statistics. (2012). *The nation's report card: Science 2011 (NCES 2012–465)*. Washington, DC: Institute of Education Sciences, U. S. Department of Education.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Needham, D. R., & Begg, I. M. (1991). Problem-oriented training promotes spontaneous analogical transfer: Memory-oriented training promotes memory for training. *Memory & Cognition*, 19, 543–557. <http://dx.doi.org/10.3758/BF03197150>
- Nokes, T. J., & VanLehn, K. (2008). Bridging principles and examples through analogy and explanation. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurow, . . . L. D'Amico (Eds.), *The proceedings of the 8th international conference for the learning sciences* (Vol. 3, pp. 100–102). Mahwah, NJ: Erlbaum.
- Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 510–520. <http://dx.doi.org/10.1037/0278-7393.14.3.510>
- Patterson, J. D., & Kurtz, K. J. (2015). Learning mode and comparison in relational category learning. In D. C. Noelle et al. (Eds.), *Proceedings of the 37th annual meeting of the cognitive science society* (pp. 1841–1846). Austin, TX: Cognitive Science Society.
- Patterson, J. D., & Kurtz, K. J. (2016). Performance pressure and comparison in relational category learning. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 2333–2338). Austin, TX: Cognitive Science Society.
- Perkins, D. (2009). *Making learning whole: How seven principles of teaching can transform education*. Hoboken, NJ: Wiley.
- Premo, J., Cavagnetto, A. R., Honke, G., & Kurtz, K. J. (2019). Categories in Conflict: Combating the application of an intuitive conception of inheritance with category construction. *Journal of Research in Science Teaching*, 56, 24–44. <http://dx.doi.org/10.1002/tea.21466>
- Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, 88, 144–161. <http://dx.doi.org/10.1037/0022-0663.88.1.144>
- R Core Team. (2015). *R: A language and environment for statistical computing* [Computer software manual]. Retrieved from <https://www.R-project.org/>
- Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, 115, 381–400. <http://dx.doi.org/10.1037/0033-2909.115.3.381>
- Rehder, B., & Ross, B. H. (2001). Abstract coherent categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 1261–1275. <http://dx.doi.org/10.1037/0278-7393.27.5.1261>
- Rittle-Johnson, B., & Star, J. R. (2007). Does comparing solution methods facilitate conceptual and procedural knowledge? an experimental study on learning to solve equations. *Journal of Educational Psychology*, 99, 561–574. <http://dx.doi.org/10.1037/0022-0663.99.3.561>
- Rittle-Johnson, B., & Star, J. R. (2009). Compared with what? the effects of different comparisons on conceptual knowledge and procedural flexibility for equation solving. *Journal of Educational Psychology*, 101, 529–544. <http://dx.doi.org/10.1037/a0014224>
- Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249–255. <http://dx.doi.org/10.1111/j.1467-9280.2006.01693.x>
- Ross, B. H. (1984). Reminders and their effects in learning a cognitive skill. *Cognitive Psychology*, 16, 371–416.
- Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 629–639. <http://dx.doi.org/10.1037/0278-7393.13.4.629>
- Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 427–440. <http://dx.doi.org/10.1037/0278-7393.23.2.427>
- Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science*, 36, 919–932. <http://dx.doi.org/10.1111/j.1551-6709.2012.01253.x>
- Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, 36, 1019–1050. <http://dx.doi.org/10.1111/j.1551-6709.2012.01250.x>
- Scheiter, K., Gerjets, P., & Schuh, J. (2010). The acquisition of problem-solving skills in mathematics: How animations can aid understanding of structural problem features and solution procedures. *Instructional Science*, 38, 487–502. <http://dx.doi.org/10.1007/s11251-009-9114-9>
- Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, 16, 475–522.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26, 559–569. <http://dx.doi.org/10.1177/0956797614567341>
- Sloutsky, V. M., Kaminski, J. A., & Heckler, A. F. (2005). The advantage of simple symbols for learning and transfer. *Psychonomic Bulletin & Review*, 12, 508–513. <http://dx.doi.org/10.3758/BF03193796>
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press. <http://dx.doi.org/10.4159/harvard.9780674866270>
- Smith, J. I., Combs, E. D., Nagami, P. H., Alto, V. M., Goh, H. G., Gourdet, M. A. A., . . . Tanner, K. D. (2013). Development of the biology card sorting task to measure conceptual expertise in biology. *CBE Life Sciences Education*, 12, 628–644. <http://dx.doi.org/10.1187/cbe.13-05-0096>

Spencer, R. M., & Weisberg, R. W. (1986). Context-dependent effects on analogical transfer. *Memory & Cognition, 14*, 442–449. <http://dx.doi.org/10.3758/BF03197019>

Su, Y. S., Gelman, A., Hill, J., & Yajima, M. (2011). Multiple imputation with diagnostics (mi) in R: Opening windows into the black box. *Journal of Statistical Software, 45*, 1–31. <http://dx.doi.org/10.18637/jss.v045.i02>

Thorndike, E. L. (1924). Mental discipline in high school studies. *Journal of Educational Psychology, 15*, 1–22. <http://dx.doi.org/10.1037/h0075386>

Trench, M., & Minervino, R. A. (2015). The role of surface similarity in analogical retrieval: Bridging the gap between the naturalistic and the experimental traditions. *Cognitive Science, 39*, 1292–1319.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 382–402). New York, NY: Academic Press.

Vendetti, M. S., Wu, A., Rowshanshad, E., Knowlton, B. J., & Holyoak, K. J. (2014). When reasoning modifies memory: Schematic assimilation triggered by analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1172–1180. <http://dx.doi.org/10.1037/a0036350>

Whitehead, A. N. (1929). *The aims of education and other essays*. New York, NY: Macmillan.

Zhang, Z., & Yuan, K. H. (2018). *Practical statistical power analysis using Webpower and R*. Granger, IN: ISDSA Press.

Appendix A

Transfer Problem

Please read the following passage carefully and thoughtfully, and then proceed to the accompanying question.

Once a virus enters your body, it infects a cell by injecting DNA into it. This viral DNA instructs the cell to produce thousands of copies of

the virus. However, no matter how many virus copies are made the immune system always recognizes them as foreign and attacks when they exit the cell.

How can the Virus Leave the Cell AND Survive?

Appendix B

Experiment 1a and 2–Solved Problem Cases

Target principle cases	Contrast cases
Carnivorous plants often emit a stench mimicking that of rotting flesh to attract their prey. Maggots are drawn towards the stench of their regular feeding ground—a dead corpse—and are then devoured by the plant.	The viceroy butterfly is a nontoxic creature that escapes its predator by displaying a pattern similar to that of the monarch butterfly, which is toxic to its predators. Its predators believe it to be a Monarch and, therefore, avoid it.
An ancient Mayan practice after a victorious battle was to remove and wear the distinctive clothing of their enemy. Dressed in the clothing of the deceased, they would enter enemy territory. From a distance, families and loved ones of the deceased believed their loved ones were returning home from battle; it was not until the Mayans were too close that their fatal mistake was realized.	Certain rare species of birds have the ability to mimic the sound of a rattlesnake. When these birds come in contact with a predator, they make this noise and the predator flees the area, mistakenly thinking that a rattlesnake is nearby.
The Queen knew to be careful, so she had her food sampled by the royal taster before she would partake. However, her enemies came to know that the Queen had an insatiable appetite for a rare type of nut. The enemies influenced Her Majesty’s favorite baker to send the Queen a precious chocolate containing the rare nut filling and also a deadly poison. The Queen joyously popped the treat in her mouth all at once. The treat would be her last.	During the war a large contingent of soldiers were stranded in an area vulnerable to an attack from the sky. With no other defense they decided to use the surrounding debris to create objects that would appear to be anti-aircraft weaponry from a plane. When a fighter plane passed over the camp, it went into evasive maneuvers and the men below were spared.

Note. Top and middle left cases used for the comparison condition.

(Appendices continue)

Appendix C

Experiment 1a and 2–Unsolved Problem Cases

Target principle cases	Contrast cases
<p>Carnivorous plants eat insects. Insects do not naturally approach these plants; instead the insects are drawn to the scent of rotting animal flesh. Some carnivorous plants all have the ability to emit chemicals into the air that mimic particular smells. How do the plants get their food?</p> <p>An ancient Mayan practice after a victorious battle was to gather the weapons and clothes of their fallen foes. The Mayans intended to conquer the enemy's villages after the battle to seize the land for their own. If the villagers were able to recognize Mayan warriors approaching, they could flee and strike back at a later time. How did the Mayans prevent this?</p> <p>The Queen knew to be careful, so she had her food sampled by the royal taster before she would partake. However, her enemies came to know that the queen had an insatiable appetite for a rare type of nut. These enemies also learned that the royal baker was sympathetic to their cause. What might the conspirators do to assassinate the queen?</p>	<p>The viceroy butterfly is a nontoxic creature that has no natural defenses. Monarch butterflies have deadly toxins and distinctive patterns that mark the toxicity and scare predators from eating them. If the viceroy cannot evolve toxins, what can it do so that predators fear it like the monarch butterfly?</p> <p>Certain rare species of birds have the ability to mimic the sounds of the predators that live in their environment. These birds have no natural defenses to counter the attacks of predators. What can the birds do to protect themselves when predators are nearby?</p> <p>During the war a large contingent of soldiers were stranded in an area vulnerable to an attack from the sky. They had no defenses to protect themselves from air attack. The soldiers did have several broken anti-aircraft weapons that would appear to be functional from the cockpit of a plane. What can the soldiers do to minimize the possibility of being attacked from the air?</p>

Note. Top and middle left cases used for the comparison condition.

Appendix D

Experiment 1b–Unsolved Problem Cases

Target principle cases	Contrast cases
<p>Carnivorous plants eat insects, but they need a way to catch their food without being able to move from their spot. The insects do not naturally approach these plants; instead the insects are drawn to the scent of rotting animal flesh. The carnivorous plants come in various sizes, shapes, and colors, but what they all have in common is the ability to emit chemicals into the air that mimic particular smells. How do the plants get their food?</p> <p>An ancient Mayan practice after a victorious battle was to gather the weapons and clothes of their fallen foes. The Mayans intended to plunder their enemy's village after the battle to seize more valuables. If the villagers were able to recognize Mayan warriors approaching, they could hide their treasures or flee with them. How did the Mayans prevent this?</p> <p>A food critic has become so well-known that she can no longer do her job. The staff at the best restaurants know her by sight. If she is spotted, the staff and chef will make every effort to impress her. Of course, the food critic is supposed to evaluate how the restaurant serves an ordinary customer, not a special one. In order to judge the restaurants fairly, what can the famous food critic do?</p>	<p>Two women who run a beachside gift shop make almost all of their money during the summer season. However, they have expenses all year long that require having funds on hand. What do they do to make sure they have enough money throughout the year despite only bringing in income during the summertime?</p> <p>Nearly all birds have the ability to fly. Flight allows birds to migrate long distances and nest in places that ground dwellers cannot reach. To fly, the birds must be relatively lightweight. Bones are the heaviest part of an organism, and ordinary bones would make flight difficult. How can birds have bones and still be able to fly?</p> <p>Cellular telephone calls require a set of conditions to be met to transmit a person's voice from one phone to another. One of the biggest reasons that calls fail is because the phone falls out of range from a tower. This can happen when the signal is blocked by mountains or large buildings. How might cellphone providers decrease the frequency of dropped calls?</p>

Note. Top and middle left cases used for the comparison condition.

(Appendices continue)

Appendix E

Experiment 3–Study Cases

Target principle cases	Contrast cases
<p>Carnivorous plants eat insects. Insects do not naturally approach these plants; instead the insects are drawn to the scent of rotting animal flesh. Some carnivorous plants all have the ability to emit chemicals into the air that mimic particular smells. How do the plants get their food?</p> <p>An ancient Mayan practice after a victorious battle was to gather the weapons and clothes of their fallen foes. The Mayans intended to conquer the enemy's villages after the battle to seize the land for their own. If the villagers were able to recognize Mayan warriors approaching, they could flee and strike back at a later time. How did the Mayans prevent this?</p> <p>The Queen knew to be careful, so she had her food sampled by the royal taster before she would partake. However, her enemies came to know that the queen had an insatiable appetite for a rare type of nut. These enemies also learned that the royal baker was sympathetic to their cause. What might the conspirators do to assassinate the queen?</p> <p>Cellular telephone calls require a set of conditions to be met in order to transmit a person's voice from one phone to another. One of the biggest reasons that calls fail is because the phone falls out of range from a tower. This can happen when the signal is blocked by mountains or large buildings. How might cellphone providers decrease the frequency of dropped calls?</p>	<p>The viceroy butterfly is a nontoxic creature that has no natural defenses. Monarch butterflies have deadly toxins and distinctive patterns that mark the toxicity and scare predators from eating them. If the viceroy cannot evolve toxins, what can it do so that predators fear it like the monarch butterfly?</p> <p>Certain rare species of birds have the ability to mimic the sounds of the predators that live in their environment. These birds have no natural defenses to counter the attacks of predators. What can the birds do to protect themselves when predators are nearby?</p> <p>During the war a large contingent of soldiers were stranded in an area vulnerable to an attack from the sky. They had no defenses to protect themselves from air attack. The soldiers did have several broken anti-aircraft weapons that would appear to be functional from the cockpit of a plane. What can the soldiers do to minimize the possibility of being attacked from the air?</p> <p>Two women who run a beachside gift shop make almost all of their money during the summer season. However, they have expenses all year long that require having funds on hand. What do they do to make sure they have enough money throughout the year despite only bringing in income during the summertime?</p>

Note. Bottom left and bottom right cases used as unrelated distractors.

Received February 21, 2019
Revision received June 11, 2019
Accepted June 17, 2019 ■