Cognition 196 (2020) 104146

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Relational labeling unlocks inert knowledge

Anja Jamrozik*, Dedre Gentner

Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208, United States of America

ARTICLE INFO

Keywords:
Analogy
Memory retrieval
Relational transfer
Relational reasoning
Creativity

ABSTRACT

Insightful solutions often come about by recalling a relevant prior situation—one that shares the same essential relational pattern as the current problem. Unfortunately, our memory retrievals often depend primarily on surface matches, rather than relational matches. For example, a person who is familiar with the idea of positive feedback in sound systems may fail to think of it in the context of global warming. We suggest that one reason for the failure of cross-domain relational retrieval is that relational information is typically encoded variably, in a context-dependent way. In contrast, the surface features of that context—such as objects, animals and characters—are encoded in a relatively stable way, and are therefore easier to retrieve across contexts. We propose that the use of relational language can serve to make situations' relational representations more uniform, thereby facilitating relational retrieval. In two studies, we find that providing relational labels for situations at encoding or at retrieval increased the likelihood of relational retrieval. In contrast, domain labels—labels that highlight situations' contextual features—did not reliably improve domain retrieval. We suggest that relational language allows people to retrieve knowledge that would otherwise remain inert and contributes to domain experts' insight.

1. Introduction

Vocabulary learning is often portrayed as a pointless exercise assigned to keep students busy. Yet learning certain kinds of terms may be extremely valuable for achieving domain mastery. The rationale for this is that higher-order cognition depends on relational processing—on the ability to represent and reason about relations such as *causation* and *prevention* in science, *commutativity* and *distributivity* in mathematics, and *promise* and *lie* in social interactions. We suggest that learning and using terms that denote these relational patterns is instrumental in acquiring cognitive flexibility and insight in and across domains.

In this paper we focus on one specific way in which language can support cognition— relational retrieval from long-term memory. Decades of research in the laboratory have shown poor retrieval of relational matches from memory. Given a current situation, people are often reminded of prior situations that share specific content features—objects, characters, locations, and associated entities—and fail to retrieve those that share relational structure (Brooks, 1987; Brooks, Norman, & Allen, 1991; Forbus, Gentner, & Law, 1995; Gentner, Rattermann, & Forbus, 1993; Holyoak & Koh, 1987; Ross, 1987, 1989; Trench & Minervino, 2015). This is true even when the relational match is demonstrably stored in memory (Gentner et al., 1993; Gick & Holyoak, 1980, 1983; Holyoak & Koh, 1987; Trench & Minervino, 2015); and it holds even when the same people later rate the relational matches (that did not come to mind) as both more similar and more inferentially sound than the surface matches that they readily retrieved (Gentner et al., 1993).

1.1. Encoding variability

One way this pattern of rare relational remindings has been interpreted is in terms of differential encoding variability between relational information and surface contextual information about the entities that occupy roles within a relational structure (Forbus et al., 1995; Gentner, Loewenstein, Thompson, & Forbus, 2009). The idea is that relational information is often encoded in a context-specific way, and is therefore variably encoded across situations. In contrast, the entities within this context—such as objects, animals, and characters—tend to be uniformly encoded (Asmuth & Gentner, 2017; Bassok, Wu, & Olseth, 1995; Forbus et al., 1995; Gentner & France, 1988; Gentner et al., 2009;). An example of this phenomenon comes from work on *verb mutability* (Gentner, 1981; Gentner & France, 1988; Reyna, 1980). The term *mutability* refers to a word's propensity to assume different meanings across varying contexts. Evidence that verbs (which typically name relations) are more mutable than concrete nouns (which typically name object or animal categories) comes from studies by Gentner and France

* Corresponding author.
E-mail addresses: a.jamrozik@gmail.com (A. Jamrozik), gentner@northwestern.edu (D. Gentner).

https://doi.org/10.1016/j.cognition.2019.104146
Received 17 November 2018; Received in revised form 19 November 2019; Accepted 20 November 2019

(1988) in which people were asked to paraphrase semantically strained sentences such as "The car worshipped." The dominant response was to use a synonym for the noun (roughly preserving the noun's usual referent) and to alter the meaning of the verb to fit that referent: e.g., "The vehicle only responded to him," or "Someone's vehicle was given a rest on a Sunday."

Verb mutability has implications for the encoding and recognition of verbs versus concrete nouns (Earles & Kersten, 2017; Kersten & Earles, 2004; King & Gentner, 2019). For example, Earles and Kersten (2017) gave people simple sentences to remember, and later tested them with different combinations of words. In the test, they were asked to recognize a specific target word—either a verb or a noun—within each test sentence. Overall, nouns were better recognized than verbs. More to the point, verbs, but not nouns, showed a significant deficit from change of context. People were worse at recognizing verbs when they were paired with new nouns than when they appeared with the original nouns—for example, if the original was "drop the spoon," people were better able to recognize "drop" in that same phrase than in a new phrase "drop the photograph". In contrast, for nouns, recognition was equally good whether they were paired with different verbs or the same verb—for example, people were equally able to recognize "spoon" in the new phrase "bend the spoon" as in the previously-seen phrase "drop the spoon" (Earles & Kersten, 2017). These findings support Gentner's (1981) verb mutability hypothesis. Because the verbs were encoded differently in the new context than in the original context, the likelihood of experiencing a match between the test sentence and the input sentence was relatively low. In contrast, the nouns were likely to be encoded in the same way in both contexts, making it likely that people would experience a match.

Forbus et al. (1995) proposed that more generally, relations tend to be encoded in a context-specific way. And because spontaneous reminding from long-term memory depends on a match between the current situation and a stored representation, elements that are uniformly encoded (such as objects and characters), have an advantage in reminding over relational patterns, which are variably encoded.

1.2. Relational retrieval and expertise

Despite much evidence demonstrating poor relational retrieval, people do sometimes retrieve purely relational matches, with little or no surface support (Gentner et al., 1993). In the world outside the lab, creative solutions in science, design, and technology often come about because of spontaneous analogies, which rely on relational mappings within and between domains (e.g., Dunbar, 1995; Hargadon & Sutton, 1997; Majchrzak, Cooper, & Neece, 2004). For instance, designers at Nike developed a shock-absorbing shoe by drawing on a solution from Formula One race car suspension (Kalogerakis, Lüthje, & Herstatt, 2010). There is evidence that relational remindings become more likely as people gain in domain knowledge (Goldwater & Schalk, 2016; Goldwater, Sibley, Gentner, LaDue, & Libarkin, under review; Koedinger & Roll, 2012; Novick, 1988). This cannot be due to simple accrual of examples, because this would also permit many more potential surface matches. A possible explanation, based on the above discussion, is that as people gain in domain knowledge, their representations become more sophisticated; they come to encode domain phenomena in terms of a set of higher-order relational schemas that apply widely in the domain (such as *positive feedback loop* or *conservation of energy*) (Forbus et al., 1995; Gentner et al., 2009; Goldwater & Gentner, 2015). The habitual use of key relational patterns during encoding promotes uniform relational representation, making it likely that a current example will overlap relationally with an example or abstraction stored in memory. Uniform relational representation could thus contribute to domain experts' superior relational retrieval.

How might experts' relational uniformity come about? There could be many contributing factors. People's self-explanations may contribute to deeper, more consistent representations of a domain (Chi, Feltovich, & Glaser, 1981; Legare & Lombrozo, 2014; Lombrozo, 2016). Analogical comparisons (whether guided or discovered) can highlight a common system of relations, inviting a more uniform representation across the two analogs (Catrambone & Holyoak, 1989; Day & Asmuth, 2017; Doumas & Hummel, 2013; Gentner et al., 2009; Gentner & Christie, 2010; Gick & Holyoak, 1983; Loewenstein, Thompson, & Gentner, 1999). We propose another factor that can contribute to uniform relational encoding—learning the relational language that characterizes a domain. We focus particularly on terms that name relational categories—for example, *carnivore* or *positive feedback loop*. Relational categories are categories whose members do not in general share common intrinsic features; rather, category membership is based on a common relational pattern (Gentner & Kurtz, 2005; Goldwater & Schalk, 2016; Kurtz, Boukrina, & Gentner, 2013; Markman & Stilwell, 2001). For example, members of the relational schema category *positive feedback loop* include the increasing audio feedback resulting from a speaker being placed too close to a microphone; increasing global temperatures resulting from the melting of polar ice; and increasing contractions during childbirth resulting from the release of oxytocin. We suggest that applying a relational schema category label such as *positive feedback loop* to an example invites structuring the example in terms of the overarching schema. Habitual use of the relational vocabulary of a domain could thus promote uniform relational representation across examples, and thereby increase relational retrieval.

This account predicts that applying a relational schema label to an example during initial encoding should increase the likelihood of later relational retrieval of that example. Less obviously, it also predicts increased relational reminding of prior relationally similar examples if a relational label is applied at *test time*. This second prediction follows from evidence that deriving a schema (by comparing examples of a relational structure) at *retrieval* time improves retrieval of prior relationally similar examples (Gentner et al., 2009; Kurtz & Loewenstein, 2007)—a phenomenon dubbed 'late abstraction'. Likewise, if relational schema labels invite a corresponding relational construal, their use should promote retrieving prior examples that were encoded with the same construal. Of course, providing labels at *both* encoding and test should be especially effective; but this outcome will not be definitive, since it could result from the labels acting as a purely lexical cue that the two passages match (because they have the same label), as well as (or instead of) a relational match.

Across two studies, we used a cued-recall paradigm to test these predictions. Participants studied one set of passages in an encoding phase. After a delay, in the test phase they received a new set of passages, and were told to write any encoding passages of which they were reminded. Each test passage had the same relational structure as one of the original passages (e.g., positive feedback loop). In addition, to capture the challenge of real-life memory retrieval, for each test passage there was another original passage from the same domain as the test passage (e.g., medicine). Because the same-domain passage shared common contextual features—objects, characters, locations, and associated entities—with the test passage, we expected it to be a potent retrieval match. Thus, for each test passage there were two potential retrieval candidates: the relational match and the domain match (Fig. 1). In the absence of labels, we expected relational retrieval to be low and domain retrieval to dominate. The predictions are that (1) in the baseline condition, with no relational labels, domain retrievals will dominate; (2) providing relational labels at encoding will improve relational retrieval; and (3) this effect will also hold if relational labels are provided at test. A further prediction, tested in Experiment 2, is (4) that relational labels will lead to greater change in retrieval patterns relative to baseline than domain labels. This follows from our earlier claim that the specific content features that would be highlighted by domain labels are already encoded similarly across situations by default. We tested predictions (1), (2) and (3) in Experiments 1 and 2, and prediction (4) in Experiment 2.

In addition, in Experiment 1, we varied the placement of the label

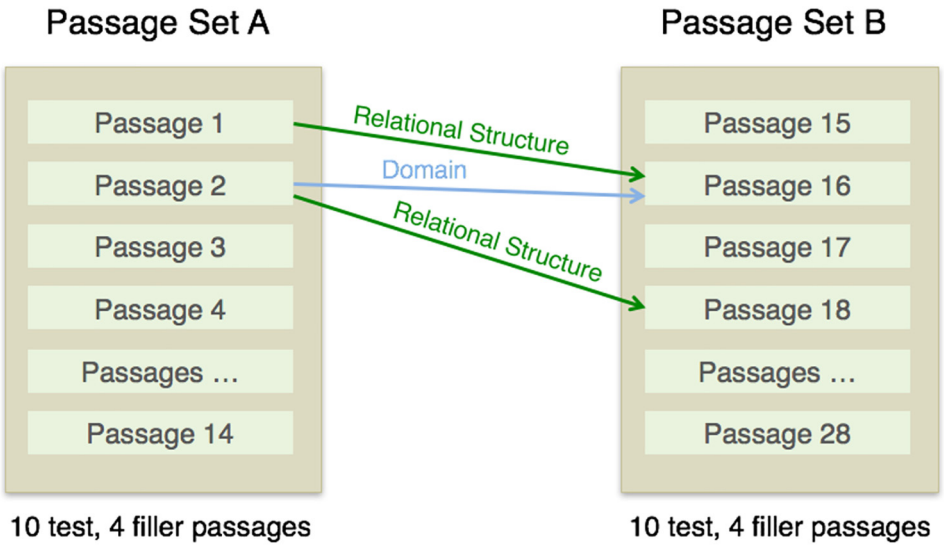

Fig. 1. The design of the encoding and test passage sets. Participants either received passage set A or B at encoding and the other set at test. Each test passage matched one encoding passage in terms of its relational structure, and a different encoding passage in terms of its domain.

within a given passage. If relational labels serve to organize the passage according to a relational pattern, then they should be more powerful if given at the start of the passage (e.g., Ausubel, 1960; Bransford & Johnson, 1972).

2. Experiment 1

2.1. Method

2.1.1. Participants

Participants (N = 251, 154 female, mean age = 19.50) were native English speakers from the Northwestern University community who were paid or received course credit for participation. An additional 15 participants were tested but excluded from analyses for failing to follow the test instructions (3), for failing to respond to at least half of the test items (10), or due to experimental error (2). The sample size of 36 participants/condition was set based on expected effect sizes from prior pilot testing with unnormed stimuli (Jamrozik & Gentner, 2013).

2.1.2. Materials and design

The materials consisted of two sets (A and B) of fourteen passages each (ten test, four filler) that served as the encoding and test sets, see Supplemental Material for the passage sets. Each passage instantiated a relational pattern that could appear across different domains. For example, the relational pattern *positive feedback loop* can appear in both atmospheric science and electrical engineering (Fig. 2). Two passages were adapted from a study by Rottman, Gentner, and Goldwater (2012).

Each passage within a set came from a different domain. Across the two sets, each test passage matched one original passage in terms of its relational pattern, and a different original passage in terms of its domain (see Fig. 1). For example, in one set, the *reciprocity* passage came from the domain of political science. In the other set, the *reciprocity* passage came from psychology, and the passage that came from political science instantiated a different relational pattern.

2.1.3. Match ratings

A separate rating task with different participants was used to verify that people perceive passages from the same domain as describing related topics (i.e., sharing specific content features like objects, characters, and locations—what the passages were about), and passages instantiating the same relational pattern as analogous (i.e., relationally similar) (see Supplemental Material for full details on these and other ratings.) Participants rated all domain match passage pairs and all relational match passage pairs on three dimensions: how *related* the passage topics were, how *analogous* (relationally similar) the passages were, and how *similar* the passages were. As expected, the relational match pairs were rated as more analogous than the domain match pairs, and the domain match pairs were rated as having higher topic-relatedness than the relational match pairs. These patterns held for all the individual passage pairs. There was also a trend suggesting that domain match pairs were rated as more similar than relational match pairs.

2.1.4. Semantic association

As an additional check on our manipulation of domain-relatedness, we used Latent Semantic Analysis (LSA) to measure the degree of semantic association between the passages (Landauer & Dumais, 1997). LSA uses patterns of word co-occurrence to quantify the degree of similarity and/or association between words or passages. For each test passage in set A, we calculated the LSA scores (Landauer & Kintsch, 1998) to its domain match and its relational match in set B. Domain match pairs had higher scores (M = 0.41, 95% CI [0.31, 0.50]) than did relational match pairs (M = 0.13, 95% CI [0.10, 0.17]), $t(9) = 7.54$, $p < .001$, confirming that, as intended, passages that came from the same domain were more semantically associated than passages that instantiated the same relational pattern.

2.1.5. Experimental design

We varied the presence of relational labels using a 2 (label present vs. absent at encoding) x 2 (label present vs. absent at test) between-subjects design. For the label conditions, an additional sentence named the relational pattern instantiated by the passage (e.g., 'This is an example of *inoculation*'). This sentence was presented either before or after the passages, varied between-subjects. This resulted in seven conditions: baseline (no labels) plus six labeling conditions: 2 (labels before/after the passage) x 3 (labels at encoding, at test, or both).

2.1.6. Label comprehensibility ratings

To verify that people understood how the relational labels related to the passages, we collected comprehensibility ratings from a separate group of participants (see Supplemental Material for details). Participants read all the test passages from either set A or set B, varied between-subjects. Half the passages had relational labels (e.g., *positive feedback loop*) and half had novel labels made up of a pseudoword

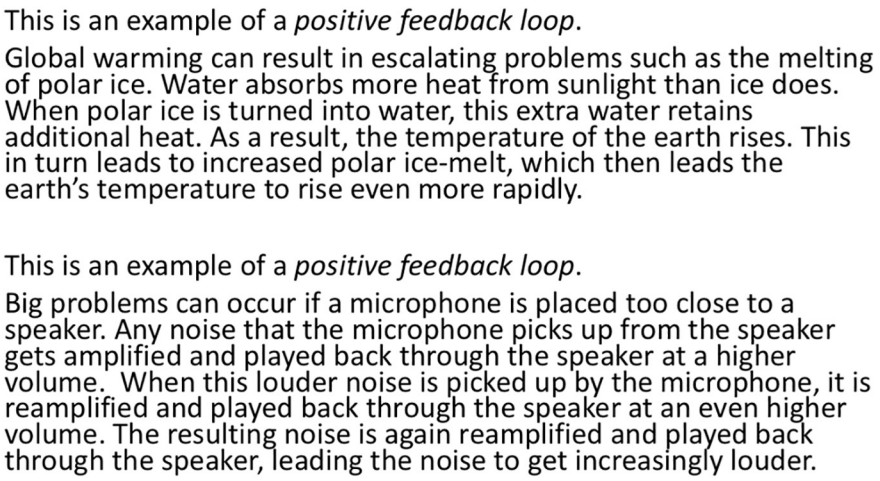

Fig. 2. An example of two test passages making up a relational match. The top passage comes from the domain of atmospheric science, while the bottom passage comes from electrical engineering.

(chosen from the ARC Nonword Database; Rastle, Harrington, & Coltheart, 2002) and a neutral relational word (e.g., *croice effect*). Participants rated how well each passage's label fit the passage and reported whether they had seen the label term used before. As expected, the relational labels were rated as more comprehensible than the novel labels, and this same pattern held for all the individual passage pairs. Relational labels were rated as comprehensible overall. Participants also reported they had seen the relational labels before.

2.2. Procedure

2.2.1. Experiment procedure

Each participant was randomly assigned to one of seven conditions: baseline (no labels), labels before/after the passages at encoding, labels before/after the passages at test, or labels before/after the passages at encoding and test. In the initial encoding phase, participants read one set of passages, either set A or B, and were told to try to remember them for a later phase of the experiment. After a 15-minute non-linguistic filler task, in the test phase, they received the other set of passages. For each passage, they were asked to write down any original passages of which they were reminded. They were told that they could write multiple original passages for each test passage and that they could write down a given original passage for more than one test passage.

2.2.2. Data coding

For each participant, we calculated the number of relational matches and the number of domain matches retrieved. Each test response was classified as *relational match*, *domain match*, or *other* by a trained research assistant blind to condition. The research assistant received all of the study materials and all of the participants' responses, grouped by the test passage they were written in response to (e.g., all participants' responses to the inoculation-medicine test passage were grouped).

For each test passage, the research assistant was told which study passage counted as a relational response and which counted as a domain response. For example, the relational match for test passage inoculation-medicine was inoculation-psychology and the domain match was trade-off-medicine. The research assistant read each of the responses and then used any keywords they contained to narrow down the passage being recalled. For example, if the response included something about cancer treatment, it would be classified as a *domain match*, because the topic of the domain match passage (trade-off-medicine) was cancer treatment and no other study passage mentioned cancer. Each passage recalled was classified as a *relational match*, *domain match*, or *other* response. If there was any doubt about which passage the response was referring to (for example, if the response could only be narrowed down to one of two options), it was marked as an *other* response.

For each participant, we calculated the number of relational matches, domain matches, and other responses for the test passages.

2.2.3. Data analysis

Since the outcomes of interest were counts (e.g., the number of relational matches retrieved), we used Poisson regression models, which are well-suited to modeling count data, to characterize the relationship between experimental condition and the number of responses retrieved by participants. The number of responses retrieved in the baseline (no-label) condition was compared to that in conditions in which participants were given relational labels. To evaluate the overall strength of labels' effect on retrieval, we calculated Bayes factors for each model compared against an intercept-only model that did not include the effect of experimental condition, using the BayesFactor package (Version 0.9.12-4.2) for R. A Bayes factor higher than one favors the model that includes experimental condition (i.e., favors the hypothesis that labels had an effect on retrieval), while a Bayes factor lower than one favors the intercept-only model. To conduct follow-up pairwise comparisons between non-baseline conditions, we used the lsmeans package (Version 2.30-0).

2.3. Results

As predicted, in the absence of labels, domain retrieval was dominant. Participants in the baseline condition retrieved over twice as many domain matches (4.58 of 10) as relational matches (1.56 of 10). However, consistent with our second prediction, the likelihood of relational retrieval increased when relational labels were provided.

Relational retrieval varied by label condition,[1] see Table 1 and Fig. 3. The model including experimental conditions was preferred over an intercept-only model (i.e., the Bayes factor is higher than 1). Participants who received labels only at encoding retrieved more relational matches than those in the baseline condition, whether the labels preceded or followed the passages. Participants who received relational labels only at test also showed this advantage over baseline, but only if the labels preceded the passages. Not surprisingly, participants who received labels at both encoding and test retrieved more relational matches than baseline participants, whether the labels preceded or

[1] The between-subjects counterbalancing variable of passage set (set A vs. set B at encoding) did not affect the number of relational matches retrieved, domain matches retrieved, total number of retrievals, nor the number of missing responses, so data from all the participants were combined in the analyses.

Table 1
Results of Poisson regression models estimating the effect of experimental condition on the number of relational matches and domain matches retrieved. The no-label condition served as the baseline for comparison. Please see the Condition estimate column for exponentiated values of summed coefficient estimates.

	B	*SE B*	*Condition estimate Exp(B intercept + B condition)*	*z*	*p*	*
Effect of relational labels on number of relational matches						
Intercept: baseline/no labels	0.4418	0.1336	1.556	3.306	.001	***
Labels at encoding (labels before passages)	0.7196	0.163	3.194	4.416	< .001	***
Labels at encoding (labels after passages)	0.7819	0.1621	3.400	4.825	< .001	***
Labels at test (labels before passages)	0.4798	0.1691	2.513	2.837	.005	**
Labels at test (labels after passages)	0.2369	0.1799	1.971	1.317	.188	
Labels at encoding & test (labels before passages)	1.2928	0.1509	5.667	8.569	< .001	***
Labels at encoding & test (labels after passages)	1.4718	0.1482	6.777	9.933	< .001	***
Bayes factor vs. intercept only model	2.453E+26 ± 0%					
Effect of relational labels on number of domain matches						
Intercept: baseline/no labels	1.522	0.078	4.583	19.556	< .001	***
Labels at encoding (labels before passages)	−0.595	0.131	2.528	−4.557	< .001	***
Labels at encoding (labels after passages)	−0.493	0.128	2.800	−3.864	< .001	***
Labels at test (labels before passages)	0.008	0.109	4.622	0.076	.939	
Labels at test (labels after passages)	−0.136	0.115	4.000	−1.185	.236	
Labels at encoding & test (labels before passages)	−0.711	0.136	2.250	−5.244	< .001	***
Labels at encoding & test (labels after passages)	−0.886	0.144	1.889	−6.151	< .001	***
Bayes factor vs. intercept only model	4.206E+10 ± 0%					

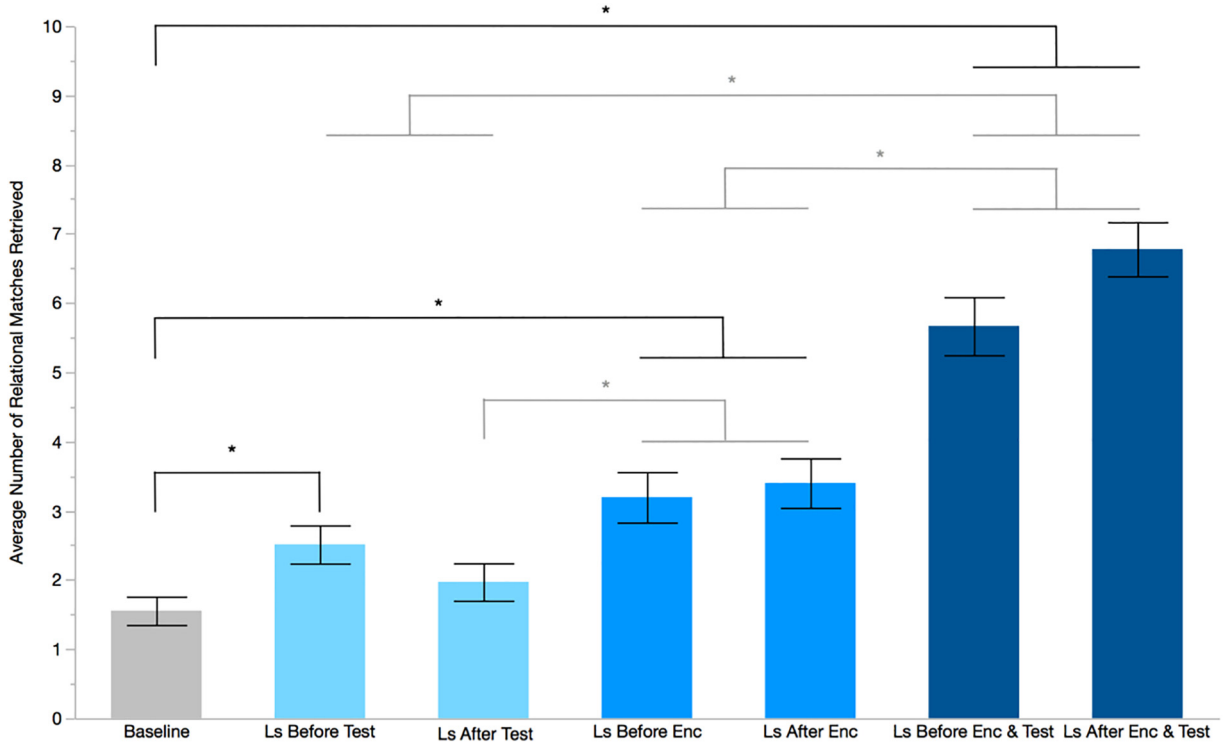

Fig. 3. Results of Experiment 1. The main comparisons of interest, marked in black, were between the baseline (no-label) condition and each of the labeling conditions. Participants who received relational labels at test (labels before passages), at encoding (labels before/after passages), or at encoding and test (labels before/after passages) all retrieved more relational matches than those who did not receive labels. Additional differences between conditions are noted in grey.

followed the passages. See Supplemental Material for descriptive statistics by condition.

Additional Tukey-adjusted pairwise comparisons between conditions revealed that there were more relational retrievals when labels were present at both encoding and test than when they were present only at test or only at encoding (regardless of whether the labels were before or after the passages), all *ps* < .001. Finally, there was an advantage for labels only at encoding (labels before or after the passages) over labels only at test (labels after passages), $p = .026$ and $p = .006$, respectively.

Relational labels also had an effect on domain retrieval, see Table 1. The model including experimental conditions was preferred over an intercept-only model. Participants who received relational labels mostly retrieved fewer domain matches. Compared to the baseline condition, fewer domain matches were retrieved by participants who received labels at encoding and test or only at encoding, regardless of label order.

Additionally, participants who received relational labels at encoding, with labels before passages, retrieved fewer domain than participants who received relational labels at test, regardless of whether test labels preceded or followed passages, $p < .001$ and $p = .012$, respectively. Likewise, participants who received relational labels at encoding, with labels after passages, retrieved fewer domain matches than participants who received relational labels at test, but only if test labels

preceded passages, $p = .002$. Participants who received relational labels at encoding and test retrieved fewer domain matches than participants who received labels at test, regardless of label order, all $ps < .001$.

3. Experiment 1 Discussion

As predicted, relational labels markedly improved the likelihood of relational retrieval. Not surprisingly, having labels at both encoding and test was highly effective: people in this condition retrieved roughly four times as many relational matches as in baseline. More importantly, receiving relational labels at encoding was also highly effective. Compared to baseline, people were twice as likely to retrieve a relational match when given a relational label at encoding. Interestingly, relational labels were also effective when given at test, though the difference from baseline reached significance only when the labels were given before the test passages.

The result that relational retrieval improved when labels were given at both encoding and retrieval fits with our predictions, in that having the same relational label in both phases invites a uniform construal of the two passages. However, we must use caution in interpreting these results, because the elevated relational retrieval could have been due in part to the relational labels acting as a purely lexical cue. In the other conditions, the label was provided for only one passage, making it impossible to match passages across phases on the basis of a label provided in-text.

Overall, Experiment 1 demonstrates that providing labels that highlight situations' relational patterns can improve people's relational retrieval. In Experiment 2, we tested the specificity of the labeling effect. We have suggested that the comparative rarity of relational matches in memory retrieval reflects the fact that relational information is encoded variably across situations, whereas information about specific content features—objects, characters, locations, and associated entities—is encoded more uniformly. This account predicts that labels that highlight these content features should not have a strong effect on retrieval patterns, because these features are already encoded similarly across situations by default. In Experiment 2, we tested this claim by comparing the effects of domain labels, which should serve to highlight examples' content features, against the effects of relational labels on retrieval.

4. Experiment 2

4.1. Method

4.1.1. Participants

Participants ($N = 114$, 80 female, mean age = 20.82) were native English speakers recruited from the Northwestern University community. They were paid or received course credit for their participation. An additional five participants were tested but excluded from analyses for failing to respond to at least half of the test items. The sample size of 17 participants/condition was based on effect sizes observed in Experiment 1. This sample size had power of 80% to detect an effect the size of that of relational labels at encoding and test vs. baseline at $p = .05$.

4.1.2. Materials, design, and procedure

The passage sets were the same as those in Experiment 1, except that the label type varied. Also, based on the results of Experiment 1, all labels were presented before the passages. The design was a 2 (label present vs. absent at encoding) × 2 (label present vs. absent at test) × 2 (relational labels vs domain labels) between-subjects design, resulting in seven conditions: baseline (no labels); relational labels at encoding, relational labels at test, relational labels at encoding and test, domain labels at encoding, domain labels at test, domain labels at encoding and test.

The procedure was as in Experiment 1. Participants in the label conditions received either relational labels (e.g., this is an example of *inoculation*) or domain labels (e.g., This is an example from *anatomy*). The dependent measures were the numbers of domain and relational matches retrieved by each participant. The same trained research assistant, blind to condition, classified responses as *relational matches*, *domain matches*, or *other* responses. As in Experiment 1, we used Poisson regression models to characterize the relationship between experimental conditions (relational labels and domain labels) on the number of responses retrieved by participants.

4.2. Results

Replicating the findings of Experiment 1, we found that relational labels affected the likelihood of relational retrieval,[2] see Table 2. The model including relational label conditions was preferred over an intercept-only model. Participants in all three of the relational label conditions retrieved more relational matches than participants in the baseline no-label condition. Tukey-adjusted pairwise comparisons did not reveal any additional differences between conditions. This pattern of results replicates the main finding of Experiment 1—that relational labels improve relational retrieval. However, surprisingly, in contrast to Experiment 1, we did not see a relational retrieval advantage for labels at *both* encoding and test over just one instance of labels.

Receiving relational labels also had the effect of reducing the rate of domain-match retrieval. As in Experiment 1, participants who received relational labels at encoding and test or only at encoding retrieved fewer domain matches than participants in the baseline condition, see Table 2. The model including relational label conditions was preferred over an intercept-only model. Tukey-adjusted pairwise comparisons did not reveal any additional differences. Thus the findings concerning the negative effect of relational labels on domain matches are parallel to those of Experiment 1, except that in Experiment 2 we did not find that participants who received relational labels at encoding or at encoding and test retrieved fewer domain matches than those who received relational labels only at test.

The pattern of results was quite different for domain labels. Consistent with the idea that information about specific content features is naturally encoded in a uniform way, domain labels did not reliably affect the likelihood of domain retrieval. An intercept-only model of domain retrieval was preferred over a model that included domain label conditions (i.e., the Bayes factor was < 1 for the model with domain label conditions as compared to the intercept-only model), see Table 2. Likewise, an intercept-only model of relational retrieval was preferred over a model that included domain label conditions, see Table 2.

Could the lack of improvement in participants' domain retrieval have been due to a ceiling effect? To examine this possibility, we used the total number of responses retrieved by participants in response to test passages as an estimate of ceiling performance. Across the two experiments, the total number of responses produced by participants to these 10 key passages was relatively consistent – between 8.6 and 10.8, and it did reliably vary by conditions in either experiment (see Supplemental Material for details). With this as the ceiling, there was still room for domain labels to have an effect. In the baseline condition in Experiment 2, domain matches made up only half of total responses to test passages. Even in the condition in which the number of domain matches was the highest (domain labels at test), this proportion was still below 60%. In comparison, in the condition with the most

[2] The between-subjects counterbalancing variable of passage set (set A vs. set B at encoding) again did not affect the number of relational matches retrieved, domain matches retrieved, total number of retrievals, nor the number of missing responses, so data from all the participants were combined in the analyses.

Table 2
Results of Poisson regression models estimating the effect of experimental condition (relational labels, domain labels) on the number of relational matches and domain matches retrieved. The no-label condition served as the baseline for comparison. Please see the Condition estimate column for exponentiated values of summed coefficient estimates.

	B	*SE B*	*Condition estimate Exp(B intercept + B condition)*	*z*	*p*	*
Effect of relational labels on the number of relational matches retrieved						
Intercept: baseline/no labels	0.894	0.151	2.444	5.929	< .001	***
Relational labels at encoding	0.573	0.189	4.333	3.037	.002	**
Relational labels at test	0.618	0.194	4.533	3.192	.001	**
Relational labels at encoding and test	0.545	0.199	4.214	2.734	.006	**
Bayes factor vs. intercept only model	2.483 ± 0.03%					
Effect of relational labels on the number of domain matches retrieved						
Intercept: baseline/no labels	1.564	0.108	4.778	14.504	< .001	***
Relational labels at encoding	−0.563	0.179	2.722	−3.143	.002	**
Relational labels at test	−0.247	0.172	3.733	−1.437	.151	
Relational labels at encoding and test	−0.442	0.187	3.072	−2.366	.018	*
Bayes factor vs. intercept only model	4.284 ± 0.02%					
Effect of domain labels on the number of domain matches retrieved						
Intercept: baseline/no labels	1.564	0.108	4.778	14.504	< .001	***
Domain labels at encoding	0.057	0.150	5.056	0.376	.707	
Domain labels at test	0.282	0.149	6.333	1.894	.058	.
Domain labels at encoding and test	0.152	0.151	5.563	1.006	.315	
Bayes factor vs. intercept only model	0.389 ± 0%					
Effect of domain labels on the number of relational matches retrieved						
Intercept: baseline/no labels	0.894	0.151	2.444	5.929	< .001	***
Domain labels at encoding	−0.201	0.225	2.000	−0.893	.372	
Domain labels at test	−0.511	0.261	1.467	−1.956	.050	.
Domain labels at encoding and test	−0.575	0.261	1.375	−2.204	.028	*
Bayes factor vs. intercept only model	0.301 ± 0%					

relational matches retrieved (relational labels after passages at encoding & test in Experiment 1), the proportion of relational matches was 70% of the total number of responses. In the baseline condition, this proportion was only 18%. This pattern of results suggests that failing to find an effect of domain labels was not due to a measurement issue. Labels *can* have a strong impact on people's retrieval patterns, as evidenced by the retrieval pattern changes brought on by relational labels. It is just that domain labels do not seem to have the same kind of effect on retrieval.

This pattern of findings is consistent with the idea that information about content features is uniformly encoded by default, and that domain labels, which serve to highlight this information, do not meaningfully change the way this information is encoded and retrieved. In sum, as predicted, relational labels had a substantially greater effect on relational retrieval than did domain labels on domain retrieval.

5. General discussion

We have advanced two claims. First, a major reason that relational retrieval is poor relative to retrieval based on surface similarity is *differential encoding variability*—information about objects and other entities is typically encoded uniformly across situations, whereas relational information is encoded in a context-specific way (Asmuth & Gentner, 2017; Forbus et al., 1995; Gentner et al., 2009; Kersten & Earles, 2004). Second, the use of a relational schema label invites a corresponding relational construal, such that a labeled example is structured to fit the overarching relational pattern conveyed by the term. This can make the representation of relational patterns across examples more uniform, thereby increasing the likelihood of relational retrieval.

Our findings support these claims. First, in both studies, in the baseline no-label condition, the number of retrieved domain matches greatly exceeds the number of relational matches, consistent with the idea that encoding of information about objects and other entities is more stable across situations than encoding of relational information. Second, the use of relational labels increased relational retrieval: in both experiments, the use of relational labels at encoding, at test, or at both encoding and test increased the likelihood of relational retrieval over a baseline no-label condition (Experiments 1 and 2). Third, this labeling effect did not occur for domain labels: domain labels, which serve to highlight information about objects, characters, and locations associated with that domain, did not reliably affect domain retrieval (Experiment 2)—consistent with the claim that this information is encoded uniformly by default. Thus, consistent with our fourth prediction, the presence of relational labels had a greater effect on relational retrieval than did the presence of domain labels on domain retrieval.

It is important to consider alternative accounts for the current set of findings. One possibility is that receiving a relational label might simply make that specific relation within the passage more salient, rather than inviting a relational construal of the larger passage. For instance, the label *positive feedback loop* might bring people's attention to this local relational pattern within a passage about global warming. Perhaps such local highlighting could help people retrieve from memory other examples that share this local pattern (e.g., a prior example about a microphone-speaker system). But if labels simply made named aspects of labeled situations more salient, then domain labels would be expected to have a similar effect by highlighting domain information and aiding retrieval of other examples sharing this information. Thus, while local salience may contribute to the effect of relational labels, we do not think it can account for the full effect. Another possibility is that people generated their own relational labels for some of the passages. If they generated a label for a passage during one phase of the experiment, receiving the same relational label for another passage in the other phase may have caused a match, aiding relational retrieval. Although self-generation is an interesting possibility to investigate, it is unlikely that this played a large role in the current studies, given the low rate of relational retrieval in the baseline condition. Finally, a third possibility is that the use of relational labels might have set an expectation that it is important, or will be important, to retrieve examples based on similarities in relational structure. This expectation would have been particularly effective if the label occurred during the encoding phase, since it would have allowed people to focus on the right details of passages to remember for later. However, again, if this account were correct, then the positive effects of expectations set by labels should have extended to

domain labels—yet they did not.

5.1. Relationship to prior work on relational and domain labels

The current findings add to prior work demonstrating that the use of relational language can improve relational retrieval and transfer. For example, Clement, Mawby, and Giles (1994) found that using the same or synonymous verbs to describe analogous situations improved relational retrieval, suggesting that using relational terms that invite a similar construal of situations can increase the likelihood of relational retrieval. Further, receiving a set of relational terms when learning about a new domain can improve the likelihood of later relational transfer in children (Loewenstein & Gentner, 2005) and adults (Son, Doumas, & Goldstone, 2010). However, to our knowledge, this research is the first to compare relational labels with domain labels. Our findings are novel in other respects as well. First, we demonstrated that adding only a *single* relational term to an example can improve relational retrieval. Second, we found that relational labels can improve retrieval even if they are present only at retrieval time. This finding extends prior research on 'late abstraction' involving comparison (Gentner et al., 2009)—suggesting that relational retrieval from memory is more likely if the probe example is encoded with a clear overarching relational pattern. We speculate that even if such an encoding makes it more likely that there will be a partial relational match with the probe example, and that this can contribute to increased relational retrieval.

In our studies, domain labels were remarkably ineffective, consistent with our claim that information about entities (e.g., objects, animals, and characters) is encoded uniformly by default. These findings dovetail with prior findings by Ripoll (1998); he found that providing a title highlighting a target example's domain did not improve retrieval of a base that shared this domain unless the base and target also shared surface cues (several matching phrases). However, we note that our findings do not imply that entity labels are psychologically unimportant across the board. There is evidence that object labels are important in forming initial nominal categories in very young children (Waxman & Hall, 1993; Waxman & Markow, 1995; Xu, 2002; Xu, Cote, & Baker, 2005) and adults (Lupyan, Rakison, & McClelland, 2007). We speculate that over repeated usage, many entity categories become strongly entrenched with a stable default representation that is readily available.

5.2. Differential encoding variability

We have suggested that one reason that relational retrieval is typically worse than retrieval via concrete entities is that relations are encoded more variably than entities. Because this is by no means a settled view, it behooves us to review evidence for this claim. As discussed earlier, one line of evidence comes from psycholinguistic studies that show a *verb mutability effect* in sentence encoding and memory (Earles & Kersten, 2017; Gentner, 1981; Gentner & France, 1988; Kersten & Earles, 2004; King & Gentner, 2019; Reyna, 1980). When people are asked to paraphrase semantically strained sentences, they are far more likely to alter the standard meaning of the verb than that of the noun (Gentner & France, 1988; see also King & Gentner, 2019; Reyna, 1980). Verb mutability is directly related to the retrieval pattern whereby verbs are poorly remembered relative to nouns, and more vulnerable to changes in context (Earles & Kersten, 2017; Kersten & Earles, 2004). As Kersten and Earles (2004, p. 199) noted: "If the meanings of verbs are dependent on linguistic context, memory for a verb may be dependent on reinstating the same linguistic context as that present when it was originally encountered."

Other evidence that relations are encoded in a more context-sensitive way than are concrete entities comes from research on how people apply mathematical operations to specific situations (Bassok, 1996; Bassok et al., 1995). For example, Bassok, Chase, and Martin (1998) asked college students to construct a simple addition (or division) problem involving two specified object categories. Despite the familiarity of these operations, the results varied substantially depending on the fit between the specified operation and the situation suggested by the pair of objects. For example, participants told to write addition problems were accurate 82% of the time for symmetrical pairs, such as tulips and daffodils, but only 61% of the time for pairs that typically occur in asymmetric situations, such as tulips and vases. Further, the most common error was to alter the specified operation to fit the content situation. For example, when given the pair peaches-basket, a participant in the addition condition wrote "Two baskets hold 30 peaches, how many peaches does 1 basket hold?" This response preserves the specified objects but alters the specified relation. Bassok and colleagues have found that for both relatively unfamiliar operations, such as permutation, and highly familiar operations, such as addition, people tend to adapt the mathematical operation to fit the entities, rather than the reverse.

5.3. Implications for education and learning

Learning terminology can seem pointless to students. However, learning the *relational* terminology of a domain may be crucial to developing expertise. Such a set can serve as a tool kit with which to uniformly encode relational patterns (Gentner, 2003, 2016; Gentner & Christie, 2010). For example, notions like 'conservation of X' or 'positive feedback loop' can be applied throughout domains. We suggest that learning and applying such relational terms helps learners to perceive and encode important relational patterns and to see non-obvious parallels within and between domains. Labeling diverse examples is important because a person who knows the meaning of a relational term in one context may not notice instances of that relational category in other contexts without some scaffolding. This is made clear by performance in the present study, in which people who did not receive relational labels were unlikely to retrieve situations that matched a target situation's relational category, even though the terms naming these relational categories are generally known.

The use of relational language may aid domain learning in other ways as well: for example, if a new term, such as *positive feedback cycle*, is applied to two examples, this invites comparing the two examples, and thereby discovering their common structure (Christie & Gentner, 2014; Gentner, 2010, 2016; Gentner & Namy, 1999). Of course, we are not arguing that learning terminology is always beneficial. Memorizing the 200 parts of a fish is unlikely to be useful for one's general understanding of biology. Rather, our findings suggest focusing on key relational terms that can be applied across situations within and across domains. We suggest that learning and using relational labels contributes to the growth of expertise. Applying a set of terms systematically throughout the domain may promote comparison and abstraction of common patterns. Over time, this habitual naming and noticing of domain-relevant relational patterns may lead to a set of highly fluent patterns that aid in interpreting domain phenomena. One reason that relational retrieval is more likely for domain experts than for novices (e.g., Goldwater et al., under review; Novick, 1988) may be that experts have acquired a technical vocabulary for common relational patterns, which they habitually use. This common vocabulary invites comparison and abstraction, contributing to the process of developing domain-general relational representations.

Indeed, as expertise increases, these patterns may become sufficiently fluent that the scaffolding provided by external labels may become unnecessary—either because of self-labeling or because labels may no longer be needed at all. In line with this idea, in two recent studies, Raynal, Clément, and Sander (2017, 2018) found that when experimental examples instantiated common relational patterns from domains that we are all expert in—social and everyday life—relational retrieval was more likely than retrieval based on superficial content features, even in the absence of overt labels. Eventually, relational categories may become well-practiced enough to not require labeling.

5.4. Implications for innovation

A relational vocabulary may help contribute to innovation in science, design, and technology, where transferring knowledge across domains is often crucial for creative solutions (Kalogerakis et al., 2010). Applying the details of previous solutions to a target problem can facilitate problem-solving and promote creativity. Examples of real life innovation (Enkel & Gassmann, 2010) illustrate that successful solutions often apply both the overall relational structure as well as some of the details of a base to a target. For instance, to solve one target problem—uneven stitches in commercial sewing—a company applied a "regulation" solution from another domain. Novice sewers have a hard time controlling the speed of sewing machines, leading to uneven stitches. What these sewers needed was a way to regulate the length of their stitches—that is, a *means of regulation* independent of sewing machine speed. The company adapted sensor technology used in computer mice that allows a mouse to track smoothly on a screen even if the speed of its physical movement is uneven. The sensor was integrated into the sewing machine foot and regulated the speed of the material moving under it to create even stitches. Using relational language to identify the problem to be solved may help people successfully search for and identify previous solutions to the same problem in order to apply them.

6. Conclusion

Our findings show that relational labels can have a powerful effect on people's ability to encode and retrieve examples of relational patterns. We suggest that the use of relational labels highlights relational patterns that might otherwise be missed, or bound to the specific features of examples. We further suggest that habitual use of domain-relevant relational terms is a major contributor to experts' superior relational encoding and retrieval.

Author contributions

A. Jamrozik: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Writing - Review & Editing.

D. Gentner: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision.

Acknowledgments

This research was funded by ONR Grant N00014-08-1-0040. We thank Laura Willig for her help in coding the data, Becky Bui, Lizabeth Huey, and Benjamin Dionysus for their administrative support, and the Cognition and Language Lab for their helpful suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.104146.

References

Asmuth, J., & Gentner, D. (2017). Relational categories are more mutable than entity categories. *The Quarterly Journal of Experimental Psychology, 70*(10), 2007–2025.
Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology, 51*, 267–272.
Bassok, M. (1996). Using content to interpret structure: Effects on analogical transfer. *Current Directions in Psychological Science, 5*(2), 54–58.
Bassok, M., Chase, V. M., & Martin, S. A. (1998). Adding apples and oranges: Alignment of semantic and formal knowledge. *Cognitive Psychology, 35*(2), 99–134.
Bassok, M., Wu, L., & Olseth, K. L. (1995). Judging a book by its cover: Interpretative effects of content on problem-solving transfer. *Memory & Cognition, 23*(3), 354–367.
Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726.
Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.). *Concepts and conceptual development: The ecological and intellectual factors in categorization* (pp. 141–174). Cambridge: Cambridge University Press.
Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120*(3), 278.
Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology Learning Memory and Cognition, 15*(6), 1147–1156.
Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121–152.
Christie, S., & Gentner, D. (2014). Language helps children succeed on a classic analogy task. *Cognitive Science, 38*(2), 383–397.
Clement, C. A., Mawby, R., & Giles, D. E. (1994). The effects of manifest relational similarity on analog retrieval. *Journal of Memory and Language, 33*(3), 396–420.
Day, S., & Asmuth, J. (2017). Re-representation in comparison and similarity. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.). *Proceedings of the 39th annual meeting of the cognitive science society* (pp. 277–282). Austin, TX: Cognitive Science Society.
Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PLoS One, 8*(6), e63889.
Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R. J. Sternberg, & J. E. Davidson (Eds.). *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.
Earles, J. L., & Kersten, A. W. (2017). Why are verbs so hard to remember? Effects of semantic context on memory for verbs and nouns. *Cognitive Science, 41*, 780–807.
Enkel, E., & Gassmann, O. (2010). Creative imitation: Exploring the case of cross-industry innovation. *R&D Management, 40*(3), 256–270.
Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19*, 141–205.
Gentner, D. (1981). Some interesting differences between verbs and nouns. *Cognition and Brain Theory, 4*, 161–178.
Gentner, D. (2003). Why we're so smart. In D. Gentner, & S. Goldin-Meadow (Eds.). *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775.
Gentner, D. (2016). Language as cognitive toolkit: How language supports relational thought. *The American Psychologist, 71*(8), 650–657.
Gentner, D., & Christie, S. (2010). Mutual bootstrapping between language and analogical processing. *Language and Cognition, 2*(2), 261–283.
Gentner, D., & France, I. M. (1988). The verb mutability effect: Studies of the combinatorial semantics of nouns and verbs. In S. L. Small, G. W. Cottrell, & M. K. Tanenhaus (Eds.). *Lexical ambiguity resolution: Perspectives from psycholinguistics, neuropsychology, and artificial intelligence* (pp. 343–382). San Mateo, CA: Kaufmann.
Gentner, D., & Kurtz, K. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.). *Categorization inside and outside the lab* (pp. 151–175). Washington, DC: APA.
Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science, 33*, 1343–1382.
Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.
Gentner, D., Rattermann, M., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology, 25*, 524–575.
Gick, M., & Holyoak, K. (1980). Analogical problem solving. *Cognitive Psychology, 12*, 306–355.
Gick, M., & Holyoak, K. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.
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.
Goldwater, M. B., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research. *Psychological Bulletin, 142*(7), 729.
Goldwater, M. B., Sibley, D. Gentner, D., LaDue, N., & Libarkin, J. (under review). Expert-novice differences in causality-based knowledge organization.
Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly,* 716–749.
Holyoak, K. J., & Koh, K. (1987). Surface and structural similarity in analogical transfer. *Memory & Cognition, 15*(4), 332–340.
Jamrozik, A., & Gentner, D. (2013). Relational labels can improve relational retrieval. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.). *Proceedings of the 35th annual conference of the cognitive science society* (pp. 651–656). Austin, TX: Cognitive Science Society.
Kalogerakis, K., Lüthje, C., & Herstatt, C. (2010). Developing innovations based on analogies: Experience from design and engineering consultants. *Journal of Product Innovation Management, 27*(3), 418–436.
Kersten, A., & Earles, J. (2004). Semantic context influences memory for verbs more than memory for nouns. *Memory & Cognition, 32*(2), 198.
King, D., & Gentner, D. (2019). Polysemy and verb mutability: Differing processes of semantic adjustment for verbs and nouns. In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.). *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 2011–2017). Montreal, QB: Cognitive Science Society.
Koedinger, K. R., & Roll, I. (2012). Learning to think: Cognitive mechanisms of knowledge

transfer. In K. J. Holyoak, & R. G. Morrison (Eds.). *The Oxford handbook of thinking and reasoning* (pp. 789–806). New York, NY: Oxford University Press (pp. 789–806).
Kurtz, K., & Loewenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition, 35*(2), 334–341.
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*(4), 1303–1310.
Landauer, T., & Kintsch, W. (1998). Latent semantic analysis. [Online computer software]. Retrieved from http://lsa.colorado.edu/.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*, 211–420.
Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. *Journal of Experimental Child Psychology, 126*, 198–212.
Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology, 50*, 315–353.
Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6*(4), 586–597.
Lombrozo, T. (2016). Explanatory preferences shape learning and inference. *Trends in Cognitive Sciences, 20*(10), 748–759.
Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science, 18*, 1077–1083.
Majchrzak, A., Cooper, L. P., & Neece, O. E. (2004). Knowledge reuse for innovation. *Management Science, 50*(2), 174–188.
Markman, A. B., & Stilwell, C. H. (2001). Role-governed categories. *Journal of Experimental & Theoretical Intelligence, 13*, 329–358.
Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology Learning Memory and Cognition, 14*(3), 510–520.
Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology, 55A*, 1339–1362.
Raynal, L., Clément, E., & Sander, E. (2017). Challenging the superficial similarities superiority account for analogical retrieval. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.). *Proceedings of the 39th annual meeting of the cognitive science society*. Cognitive Science Society: Austin, TX.
Raynal, L., Clément, E., & Sander, E. (2018). Structural similarity superiority in a free-recall reminding paradigm. In C. Kalish, M. Rau, T. Rogers, & J. Zhu (Eds.). *Proceedings of the 40th annual meeting of the cognitive science society*. Cognitive Science Society: Austin, TX.
Reyna, V. F. (1980). When words collide: Interpretation of selectionally opposed nouns and verbs. *Symposium on Metaphor and Thought*. Chicago, IL: University of Chicago, Center for Cognitive Science.
Ripoll, T. (1998). Why this makes me think of that. *Thinking & Reasoning, 4*(1), 15–43.
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*(4), 629.
Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology Learning Memory and Cognition, 15*(3), 456.
Rottman, B. M., Gentner, D., & Goldwater, M. B. (2012). Causal systems categories: Differences in novice and expert categorization of causal phenomena. *Cognitive Science, 36*(5), 919–932. https://doi.org/10.1111/j.1551-6709.2012.01253.x.
Son, J. Y., Doumas, L. A. A., & Goldstone, R. L. (2010). When do words promote analogical transfer? *The Journal of Problem Solving, 3*(1), 4.
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*(6), 1292–1319.
Waxman, S. R., & Hall, D. G. (1993). The development of a linkage between count nouns and object categories: Evidence from fifteen- to twenty-month-old infants. *Child Development, 64*, 1224–1241.
Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology, 29*, 257–302.
Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. *Cognition, 85*(3), 223–250.
Xu, F., Cote, M., & Baker, A. (2005). Labeling guides object individuation in 12-month-old infants. *Psychological Science, 16*(5), 372–377.