The Relational Luring Effect: Retrieval of Relational Information during Associative Recognition

Vencislav Popov[1], Penka Hristova[2] and Royce Anders[3]

(Accepted for publication in Journal of Experimental Psychology: General)

Here we argue that semantic relations (e.g., *works in*: NURSE HOSPITAL) have abstract independent representations in long-term memory and that the same representation is accessed by all exemplars of a specific relation. We present evidence from two associative recognition experiments that uncovered a novel *relational luring effect* (RLE) in recognition memory. Participants studied word pairs, and then discriminated between intact (old) pairs and recombined lures. In the first experiment participants responded more slowly to lures that were relationally similar (TABLE CLOTH) to studied pairs (FLOOR CARPET), in contrast to relationally dissimilar lures (PIPE WATER). Experiment 2 extended the RLE by showing a continuous effect of relational lure strength on both RTs, false alarms and hits. It employed a continuous pair recognition task, where each recombined lure or target could be preceded by 0, 1, 2, 3 or 4 different exemplars of the same relation. RTs and false alarms increased linearly with the number of different previously seen relationally similar pairs. Moreover, more typical exemplars of a given relation lead to a stronger RLE. Finally, hits for intact pairs also rose with the number of previously studied different relational instances. These results suggest that semantic relations exist as independent representations in LTM, and that during associative recognition these representations can be a spurious source of familiarity. We discuss the implications of the RLE for current models of semantic and episodic memory, unitization in associative recognition, analogical reasoning and retrieval, as well as constructive memory research.

Keywords: associative recognition; episodic memory; false memory; relations; analogical reasoning

One core-feature of human cognition is analogical reasoning. It allows us to abstract general structures away from varying experiences, and to use these abstractions to guide perception, comprehension, reasoning, and decision-making (Gentner, 2010; Gentner & Smith, 2012; Hofstadter, 2001; Hofstadter & Sander, 2013; Holyoak, Gentner, & Kokinov, 2001). While researchers have uncovered many principles that govern the mapping, abstraction, and transfer between relational structures (Gentner & Smith, 2012; Holyoak, 2012), we know less about the nature of the representation and retrieval of relational information from long-term memory (LTM). There is no consensus whether semantic relations exist as independent abstract representations in LTM, just as entity concepts do (Anderson & Lebiere, 1998; Doumas, Hummel, & Sandhofer, 2008; Estes, 2003; Estes & Jones, 2006; Gentner, 1983; Hummel & Holyoak, 1997; Kokinov & Petrov, 2001), or whether they are represented only through their specific instances and different exemplars (Collins & Loftus, 1975; Collins & Quillian, 1969; Gagné, 2001; Gagné & Shoben, 1997; Gagné, Spalding, & Ji, 2005; Leech, Mareschal, & Cooper, 2008).

Given that being able to represent, recognize, and manipulate semantic relations is fundamental for success in analogical reasoning, it is important to understand how people represent and retrieve such relations from LTM. In pursuit of this question, this paper introduces a novel effect in recognition memory, namely, a *relational luring effect* (RLE), and characterizes its implications for formal models of memory and analogical reasoning. We first turn to a discussion of the relevant literature.

1 Department of Psychology, Carnegie Mellon University, Pittsburgh, PA; 2 Department of Cognitive Science and Psychology, New Bulgarian University, Sofia, Bulgaria; 3 Aix Marseille University, CNRS, LPC, Marseille, France

We are grateful to Viktoria Atanasova, Anelia Foteva, Emi Kirilova, Kaloyan Nechev Aleksandra Ivanova and Teodor Dajdgekov for gathering the data, and to Lynne Reder for discussions about the experimental design. We appreciate the comments by Lynne Reder and Marc Coutanche on an earlier draft. We thank the anonymous reviewers whose comments and suggestions helped us to significantly improve the clarity and quality of this manuscript.

VP was supported by a Presidential Fellowship in the Life Sciences from the Richard King Mellon Foundation and the Henry L. Hillman Endowed Presidential Fellowship Fund. Preliminary results from Experiment 1 were presented as a poster during the Psychonomic Society's 57th Annual Meeting in Chicago, IL, USA.

Correspondence concerning this article should be sent to Vencislav Popov at vencislav.popov@gmail.com

How are semantic relations represented in LTM?

The importance of whether semantic relations have independent representations cannot be understated. Successful analogical reasoning, which permeates human thinking, depends fundamentally on the ability to represent, recognize, and manipulate semantic relations in an abstract manner (Doumas et al., 2008; Hummel et al., 2004). For example, without abstract representations of relations in LTM, recognition of the fact that the phrases CONCRETE WALL and GLASS WINDOW are both instances of the "*made of"* relation, would have to depend entirely on resource demanding comparison and computation within a working memory system. Yet, such relational thinking is ubiquitous (Hofstadter & Sander, 2013), can happen unintentionally (Day & Gentner, 2007; Popov & Hristova, 2015), unconsciously (Reber, Luechinger, Boesiger, & Henke, 2014), rapidly, (Estes & Jones, 2006, 2009), without involving working memory (Popov & Hristova, 2015), arises early in development (Ferry, Hespos & Gentner, 2015; Silvey, Gentner & Goldin-Meadow, 2017), and is a strong predictor of analogical ability later in life (Silvey, Gentner & Goldin-Meadow, 2017). In general, a memory system that does not allow for the abstract representation of relations, independently of their instances, would be hard-pressed to support any syntactic cognitive abilities such as language and abstract reasoning in a computationally efficient manner (Feldman, 2013; Hummel et al., 2004; Jackendoff, 2002; Roskies, 1999).

Despite this importance of relational information to human cognition, the representation of specific semantic relations is an issue overlooked by many models of LTM, which instead focus primarily on single items and entity concepts, or on novel episodic associations between them (Collins & Loftus, 1975; Collins & Quillian, 1969; Howard & Kahana, 2002; Polyn, Norman, & Kahana, 2009; Raaijmakers & Shiffrin, 1981; Reder et al., 2000; Shiffrin & Steyvers, 1997; Smith, Shoben, & Rips, 1974). These models either provide limited details about the mechanics underlying preexisting semantic relations among entities, or they delegate a secondary role for those relations.

When it comes to semantic memory, consider for instance the classic Collins and Quillian (Collins & Quillian, 1969) hierarchical network model. In this model, separate nodes represent entity concepts, while the relations between those entities are represented as labeled directional connections from one node to another (Figure 1). Here for example, semantic knowledge such as the SUN *is a type of* STAR, and STAR *is a type of* CELESTIAL OBJECT, is represented as three nodes for SUN, STAR and CELESTIAN OBJECT, and two 'arrows' that hierarchically connect them. These arrows represent an "*is a type of"* relation. However, while these connections form hierarchical classification links, it is important to clarify that relations such as "*is a type of"* do not have abstract central representations in these models, as entity concepts do, but are instead bounded to the entities that instantiate them, and are represented separately for each of their instances.

While the Collins and Quillian (1969) model has been heavily criticized and challenged by several empirical findings (Conrad, 1972; Glass, Holyoak, & Kiger, 1979; Rips, Shoben, & Smith, 1973; Rosch, 1999), many subsequent models of semantic memory continued to lack abstract central representations of relations. The updated semantic network model by Collins and Loftus (1975) discarded the hierarchical structure assumption, but relations were still represented as local links between entity nodes. In Smith, Shoben, & Rips' (1974) feature-based model, relations no longer connected entities, but entities were instead represented as sets of semantic features, which also included semantic relations. However, relations were still only stored locally within an entity's featural representation, and had no independent representations themselves.

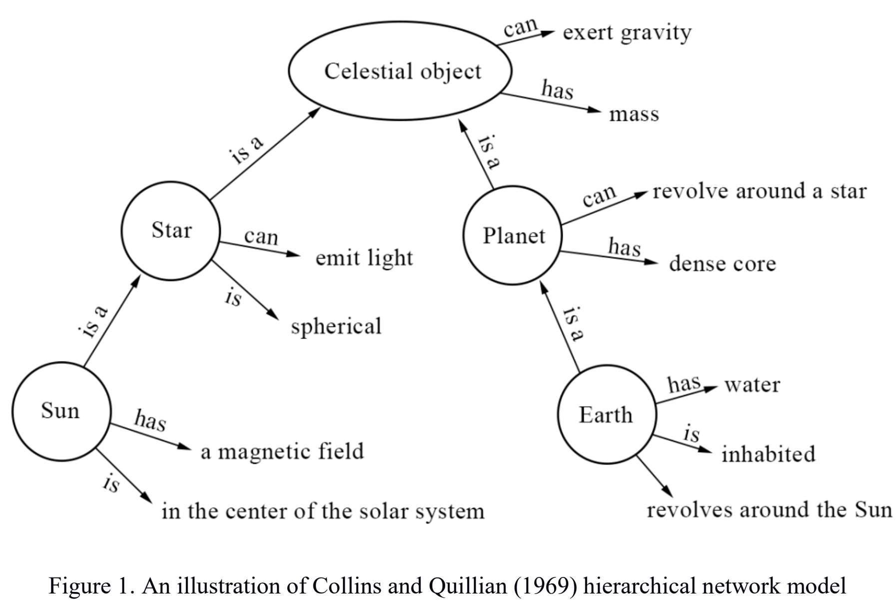

Figure 1. An illustration of Collins and Quillian (1969) hierarchical network model

The way these early models treat relational concepts is understandable, given that at that time there were no empirical studies about the nature of relations in memory and that analogical reasoning research was in its infancy. However, these models were developed to extend accounts of memory beyond merely memorizing facts into the domains of problem-solving and deductive reasoning (Tulving, 1972), and as such they should be able to account for empirical findings about the representation and retrieval of relations, such as the ones presented in this paper.

In a similar vein, more recent models of episodic memory (e.g. TCM: Howard & Kahana, 2002; CMRM: Polyn, Norman, & Kahana, 2009; SAM: Raaijmakers & Shiffrin, 1981; SAC: Reder et al., 2000; REM: Shiffrin & Steyvers, 1997) largely ignore questions about the representation and retrieval of *specific semantic relations*. Instead, they are mostly concerned with the *novel general associative links* that are created between entity items during an experiment, or with the links created between entity items and the experimental context (Buchler, Light, & Reder, 2008). In the rare cases where they do consider preexisting relations between entities, empirical tests of these models focus on general associative strength, semantic similarity or co-occurrence, while ignoring the specific relation these measures might reflect. Episodic memory models vary in the specific mechanisms they employ for single item and associative recall and recognition, but they all overlook the mnemonic nature of preexisting semantic relations.

While these types of models have remained prominent in current theories about semantic and episodic memory, they also provide a good opportunity to demonstrate how current theories of memory fail to sufficiently account for relational dynamics or analogical reasoning. As we discussed earlier, if semantic relations such as "*is a type of"* or "*is an instrument of"* do not have central abstract representations in memory, then how are individuals able to recognize the relational similarity of those two different instances? These kinds of comparisons are crucial for success in analogical reasoning. Consequently, most major models of analogical reasoning, such as SME (Falkenhainer, Forbus, & Gentner, 1989), MAC/FAC (Forbus, Gentner, & Law, 1995), LISA (Hummel & Holyoak, 1997), DORA (Doumas et al., 2008), and AMBR (Kokinov & Petrov, 2001), explicitly model relations as also having central abstract representations, just as entity concepts do (for reviews of computational models of analogy, see Doumas & Hummel, 2012, or Gentner & Forbus, 2011).

In one such model, the Structure-Mapping Theory (Gentner, 1983), relations are represented as predicates that take entity concepts or other predicates as arguments. For example, a relation such as *"revolves around"* has an abstract representation as a predicate, *revolves-around*(AGENT, PATIENT), and it can take a variety of arguments to represent any specific instance of that relation: for example, *revolves-around*(PLANET, SUN) or *revolves-around*(ELECTRON, NUCLEUS). A similar propositional notation is used by ACT-R models (Anderson & Lebiere, 1998). In addition, models like AMBR (Kokinov & Petrov, 2001) and LISA (Hummel & Holyoak, 1997) combine both propositional notation and distributed representations, where both predicates and objects are represented as distributed patterns of activation that are dynamically bound together into propositional structures. A related approach in non-analogy models is present in Rogers & McClelland's (2004) distributed semantic network. In their model, which is a variant of a model by Rumelhart & Todd (1993), relations have an input layer, which is a separate layer from that of entity items, and then relations and items are combined by converging activations into a common hidden layer. Finally, both DORA (Doumas et al., 2008) and LISA (Hummel & Holyoak, 1997) explicitly refer to dynamic role-binding as the mechanism of relational encoding and assume that the representation of relational information in LTM should be explicit and abstract so that relations can take new arguments and interact with other instances of the same relation.

In summary, the theoretical discussion so far raises the following major question when it comes to the nature of memory representations – do semantic relations have independent abstract representations in LTM, or are they represented only locally through their specific instances and different exemplars?

The relational retrieval gap

In addition to the issue of how relations are represented in LTM, another question concerns how these relations are retrieved from LTM (Holyoak, 2012). In order for analogical reasoning to occur, people first have to be able to retrieve a suitable scenario from memory that is relationally similar to the problem at hand. While some computational models of analogical reasoning do address how relational retrieval is achieved (e.g. spreading activation to relational predicate units in LISA or to relational "agents" in AMBR), most semantic and episodic memory models discussed above either do not, or only provide limited mechanisms.

Semantic network models that represent relations as connections instead of nodes (e.g. Collins & Loftus, 1975; Collins & Quillian, 1969), propose that a relation between two entities is retrieved if separate spreading activation processes initiated from each entity meet within an certain time window. However, because every instance of a relation is stored as a separate connection, such network models have no way of accounting for short-term relational priming (e.g. Estes & Jones, 2006; Popov & Hristova, 2015) and structural priming (e.g. Bock, 1986; Pickering & Ferreira, 2008; Popov & Hristova, 2014), nor for the retrieval of relationally similar structures from LTM. Finally, while recent connectionist models that represent relations as transformations between items (Leech et al., 2008) are able to account for short-term priming results, they offer no detailed account for how complex relational analogues are retrieved from LTM (French, 2008).

This latter type of relational retrieval is notoriously difficult, and has been called the "retrieval gap" (Holyoak, 2012) . Specifically, individuals often fail to spontaneously retrieve analogues from memory that share consistent relational similarities with the problem at hand, unless they are semantically similar as well (Keane, 1987), or are given hints that such analogues might be helpful for solving the current problem (Gick & Holyoak, 1980; Spellman, Holyoak, & Morrison, 2001). However, there is accumulating evidence that people's reasoning could be in fact *unintentionally* influenced by relationally similar information in memory during story comprehension (Blanchette & Dunbar, 2002; Day & Gentner, 2007; Perrott, Genter, & Bodenhausen, 2005), thematic role assignment in ambiguous sentences (Popov & Hristova, 2014), problem-solving (Day & Goldstone, 2011; Dixon & Dohn, 2003), and the comprehension of related word pairs (Estes, 2003; Estes & Jones, 2006; Popov & Hristova, 2015). These findings demonstrate that relational information can influence a variety of cognitive processes, but they also raise important questions about the precise mechanisms through which relations can be retrieved.

One major line of evidence that has been used to address these questions concerns *relational priming*. Relational priming occurs when the processing of one relational instance (e.g., the word pair planet core) influences the processing of a subsequent different, but relationally similar, instance (e.g., fruit pit; Bassok, Pedigo, & Oskarsson, 2008; Estes & Jones, 2006; Gagné, 2001; DeWolf, Son, Bassok, Holyoak, 2017; Hristova, 2009; Popov & Hristova, 2015; Raffray, Pickering, & Branigan, 2007; Spellman et al., 2001; Wisniewski & Love, 1998). Early efforts to find *unintentional* priming of relations between pairs of words were unsuccessful (Spellman et al., 2001). In addition, relational priming seemed to be restricted only to cases where there was either a repeated word across the words pairs (Gagné, 2001) or high semantic similarity between the words in the different pairs (Gagné, 2002; Gagné et al., 2005). These results lead Gagné (2005) to conclude that relations do not have independent representations that can be retrieved by different instances. However, subsequent studies that implemented better control and pretesting of the stimuli have found that relational priming occurs even when semantic similarity is controlled for (Estes & Jones, 2006; Popov & Hristova, 2015), and that it occurs unintentionally and without involving working memory resources (Popov & Hristova, 2015). On the basis of findings like these, some researchers have argued that semantic relations must be represented independently of their specific instances (Estes & Jones, 2006).

Relational priming, however, does not present sufficiently specific evidence for that conclusion. A limiting factor in relational priming research is that the relationally similar base pair is always presented immediately before the target pair. Given the short delay and the lack of intervening items between the prime and the target, we cannot know whether relational priming reflects processes that are specific to short-term memory, rather than LTM. Thus, conventional relational priming cannot be used on its own to investigate the nature of relational representations in LTM. This makes it difficult to generalize relational priming results to analogical reasoning and does not address the "retrieval gap", because the relation in these cases is already active in STM (Holyoak, 2012). In contrast, most analogical reasoning involves retrieval of analogues from LTM. Thus, we need to determine if effects similar to short-term relational priming also hold for the representations of relational information in LTM and the retrieval processes that operate on them.

In fact, there is some preliminary evidence that preexisting semantic relations can affect long-term recognition memory. For example, semantically related word pairs (e.g. HORSE RIDER) exhibit both more hits and more false alarms than non-related word pairs (DUCK BANK) in associative recognition tasks (e.g., Ahmad, Fernandes, & Hockley, 2015) and they also lead to better cued-recall (e.g., Badham, Estes, & Maylor, 2012). However, these results are consistent not only with the idea that relations have abstract representations in LTM. Even if the semantic relation between HORSE and RIDER is tied to the modifier (HORSE), as Gagné & Shoben (1997) have suggested, it could still influence memory for that specific pair. Another related finding is that when participants study a word pair such as COOKIE JAR, and are later asked to recognize the modifier word (COOKIE), they do better when it is embedded in a relationally similar word pair (COOKIE plate) compared to a relationally dissimilar pair (COOKIE crumb; Jones, Estes, & Marsh, 2008). While this is an encouraging result, it also has some of the same limitations we already discussed. Specifically, given that the experiment tested only item recognition memory for a repeated modifier, and not associative memory for novel exemplars, it is not clear whether relations in it indeed had independent representations, as the authors argued, or whether they were represented as part of the repeated modifier's representation. In addition, recent research suggests that relational nouns are encoded differently from entity nouns during memory tasks and that their semantic meaning is much more dependent on the encoding context and the entity with which they are paired (Asmuth & Gentner, 2016). Thus, whether relations are encoded abstractly in LTM, such that the same abstract representation could be retrieved by different exemplars, remains an open question.

Current experiments: The Relational Luring Effect

These questions can possibly be addressed using an associative recognition task. Previous studies have revealed that associative recognition times (RTs) and accuracy are influenced by factors such as the familiarity of each item (for a review, see Reder, Paynter, Diana, Ngiam & Dickison. 2007), the number of study repetitions of the pair (e.g., Challis & Sidhu, 1993; Reder et al., 2000), whether foils share items with previously studied pairs (e.g., Zhang, Walsh & Anderson, 2017; Asmuth & Gentner, 2016), whether the shared item in the foil is an entity or a relational noun (Asmuth & Gentner, 2016), and the degree of semantic similarity between foils and targets (Montefinese, Zannino & Ambrosini, 2015; Reagh & Yassa, 2014). Do similar effects hold for the implicit semantic relation (e.g. "*is created by*") between semantically related items in a pair (e.g., book writer)? That is, if the semantic relation is repeated multiple times through different exemplars (e.g. the exemplars book writer, blueprint architect, painting artist) would that influence reaction times and responses to novel exemplars of the relation (e.g., song composer)? If that were the case, would presenting more exemplars lead to greater effects on associative recognition decision for novel exemplars?

We propose that if semantic relations have independent abstract representations in LTM, then we should observe such effects in associative recognition. From previous research on relational priming, we know that implicit semantic relations are activated during word pair comprehension (Estes & Jones, 2006; Popov & Hristova, 2015). In addition, we propose that when participants study semantically related word pairs for a later memory test, they encode not only the individual items, but also the implicit semantic relation between them. If these relations are represented abstractly in long-term memory, then every time a participant sees a different exemplar of a relation, the activation strength of that relation in LTM should increase. In turn, it is likely that during testing, recognition decisions are based not only on retrieving the individual items and their episodic association, but also on retrieving the semantic relation between them. Since recognition memory is strongly influenced by the activation strength of items in LTM (Reder et al., 2000), the activated relation might be cause spurious familiarity when responding to other exemplars.

We tested this *relational luring hypothesis* with two associative recognition experiments, where participants studied a number of word pairs and were tested on intact and recombined pairs of the studied words. Crucially, the words in some of the test pairs (e.g. FLOOR CARPET) had the same relation (*is covered by)* as the words in some of the studied word pairs (TABLE CLOTH). If the relational luring hypothesis is correct, then the shared relation may serve as a false source of familiarity during recognition, which should lead to increased false alarm rates and to slower response times for lure pairs that are relationally similar to studied pairs. In Experiment 1 the recombined pairs shared their relation with either 0 (non-relational lures) or 1 (relational lures) of the study pairs. Experiment 2 extended the results by manipulating the strength of the relational lures in a continuous pair recognition task, where new, recombined, and intact pairs were intermixed into a single list without separate study and test sessions. Both recombined lures or intact pairs in Experiment 2 were preceded by 0, 1, 2, 3 or 4 different exemplars of the same relation (0 = non-relational lures, 1-4 = relational lures with increasing strength).

Experiment 1

Method

Participants. Forty undergraduate students (12 males) at New Bulgarian University participated for partial fulfillment of course credit. All were native Bulgarian speakers, whose age ranged from 18 to 32 years (M = 22.4, SD = 3.7). The study was approved by the Department of Psychology Ethics Committee.

Procedure. There were three blocks of stimuli and each consisted of a study phase, a distracter phase, and a recognition phase. During each study phase, participants saw 21 word pairs presented individually on a computer screen. Each study trial began with a fixation cross presented for 1 second (s.), followed by a word pair presented in the middle of the screen for 4 s., one word above the other. Then an empty screen appeared for 500 milliseconds (ms.) at the end of each trial. Prior to the study phase, participants were instructed to read and remember the word pairs, as they would have to recognize them on a subsequent test.

Table 1 A single set of items in Exp 1

ID	Pair to be tested	Relation
X	floor carpet	is covered by
Y	table cloth	is covered by
A	pipe water	flows through
B	artery blood	flows through

After the last study trial in each block, participants performed a distracter task for 1 minute, where they had to count backwards from 60 by threes - 60, 57, 54, etc. A recognition test immediately followed the counting task. As in the study phase, test pairs were presented one at a time on the computer screen. There were 21 pairs – 7 that have been studied, 7 recombined relational lures (word pairs that were relationally similar to one of the studied word pairs), and 7 recombined non-relational lures (word pairs that were not relationally similar to any of the studied pairs). Only words from the study phase were used in the test phase, and their relative position with respect to the fixation point, was preserved even for foils (lures). As for the response, participants had to press one button if they had previously studied these two words in the same pair during the study phase (“intact” response), or another button if they think the words had been studied in separate pairs during the study phase (“recombined” response). Participants were told to respond as quickly and as accurately as possible. If they failed to respond within 4 s, the current trial was terminated and then the next one commenced.

This procedure was repeated 3 times for a total of 63 studied word pairs. Word pairs were presented in a different random order for each participant in both the study and the test phase. There was a training phase consisting of 12 trials in the beginning of the experiment that consistent of random pairings of words, which were not used in the rest of the experiment.

Materials and counterbalancing. We took 84 word pairs from the pool of items used by Popov & Hristova (2015). Each word pair was matched with one other word pair so that the words in the two word pairs were related in the same way. For example, the word pair [FLOOR CARPET] was matched with the word pair [TABLE CLOTH], because FLOOR relates to CARPET in the same way as TABLE relates to CLOTH (e.g., the CARPET/CLOTH *covers* the FLOOR/TABLE). These word pairs were

Table 2 Studied and tested pairs from a single set for 2 different participants.

Participant	ID	Studied pair	Tested pair	Condition
1	X	floor carpet	floor carpet	intact
1	Y	table water	table cloth	recombined relational lure
1	A	pipe cloth	pipe water	recombined non-relational lure
2	A	pipe water	pipe water	intact
2	B	artery cloth	artery blood	recombined relational lure
2	Y	table blood	table cloth	recombined non-relational lure

controlled for relational similarity, co-occurrence, and semantic similarity; and the individual words in each pair were controlled for length, written frequency, and orthographic neighborhood (For more details, see Popov & Hristova, 2015). It is important to note that the stimuli were in Bulgarian, and in contrast to English, noun-noun combinations are not grammatical phrases in Bulgarian, and *cannot be perceived as a single unit*.

We counterbalanced the stimuli so that each word pair was randomly permuted to be in one of the three conditions (intact, relational lures, non-relation lures) for each participant with the constraint that across participants, each pair appeared equally often in each condition. To achieve this, we grouped the stimuli into 21 sets of 4 word pairs. A single set is presented in Table 1. Specifically, in each set of 4 word pairs, two of the word pairs were similarly related (as in the example above) to one another, and the other two pairs were similarly related yet in a different respect. However, between the two subsets the pairs were not relationally similar to one another. For example, for the pairs presented in Table 1, X is relationally similar to Y, and A is relationally similar to B, but all other combinations are not relationally similar.

Each participant was tested on only three of the four word pairs in each set (for example, X, Y and A, but not B), and these were rotated over participants. In this way, across participants each word pair was used in each condition (intact/relational lure/non-relational lure). For the study phase, one of the word pairs was presented intact while the words from the other two were recombined. An example for two different participants using the single set from Table 1 is presented in Table 2.

Design. The study used a within-subject design with a single factor, specifically, the type of word pair presented during the recognition phase ("intact", "recombined relational lure", "recombined non-relational lure").

Results

We used *R* (R Core Team, 2014) and *lme4* (Bates, Mächler, Bolker, & Walker, 2014) to analyze the accuracy data via logistic mixed effects regressions and response times via linear mixed effects regressions, both with participants and items as random intercept effects (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). We excluded incorrect responses from the analysis of the response times (17%). There were no random slopes that improved any of the models. Response times were log transformed, because the residual plots revealed a lack of homoscedasticity (there was no difference in the regression-analysis conclusions between analyzing the raw and the log transformed response times). All p values were obtained by likelihood ratio tests of the regression model that compared the effect in question to an identical regression model that lacked only this effect. Furthermore, factors were added step-wise to the regression models in the order they are reported.

Response times. We controlled for practice or fatigue effects before looking at the effect of interest by including block and trial position as regressors in our analysis. Response times became faster with each block (ΔAIC = -5, LLR χ^2 (1) = 7.539, $p < .01$) and with each trial within the blocks (ΔAIC = -42, LLR χ^2 (1) = 43.927, $p < .001$). The effect of trial number was not linear, but instead followed a power law – response times quickly decreased after the first few items, but then leveled-off (Newell & Rosenbloom, 1981). Indeed using the inverse trial number (rather than trial number) as an effect, improved the model (ΔAIC = -78, LLR χ^2 (0) = 78.25, $p < .001$).

The key result was that participants responded most slowly to relational lures (M = 1579, SD = 574), followed by non-relational lures (M = 1519 ms., SD = 550 ms.), which in turn were slower than intact pairs (M = 1384 ms., SD = 526 ms.; ΔAIC = -95, LLR χ^2 (2) =98.122, $p < .001$). Importantly, the contrast in response times between relational and non-relational lures (60 ms., 95%CI: 14-104 ms.) was significant ($z = 2.45$, $p < .05$). In summary, as we hypothesized participants responded more slowly when the lure was relationally similar to one of the studied pairs.

To explore this result in more detail, we split the data based on whether the relational lure (e.g., [FLOOR CARPET]) was presented before or after its corresponding intact pair (e.g., [TABLE CLOTH], Figure 2). The former case ('REL before INTACT' condition) is more interesting, because any effect found here must be due to the spontaneous retrieval of relational information. In the latter case ("REL after INTACT' condition), however, an effect might be due to the fact that the relation is still active in working memory. We included the pair presentation order in the regression model of RTs as a factor, and notably as a possible interaction with the test pair type. We found a significant interaction between the pair presentation order and test pair type (ΔAIC = -2, LLR χ^2 (2) = 5.992, $p < .05$), shown in the bottom panel of Figure 2. Post-hoc tests revealed that the difference in response times between relational and non-relational lures was present only when relational lures were presented before the corresponding intact pair, (z = 2.336, p < .05), but not when it was presented after it (z = 1.535, p = .12).

False alarm rates. False alarm rates decreased with each trial (ΔAIC = -24, LLR χ^2 (1) = 16.165, p < .001), but we did not find differences across the three lists (ΔAIC = 0.1, LLR χ^2 (2) = 3.950, p = .14). Surprisingly, there was no difference in the false alarm rates between the relational lures (M = 0.19, SD = 0.39) and non-relational lures (M = 0.17, SD = 0.37; ΔAIC = 0.3, LLR χ^2 (1) = 1.701, p = .19), nor was there an effect of the presentation order during recognition (ΔAIC = 1.7, LLR χ^2 (1) = 0.306, p = .58). The interaction between type of lure and presentation order was not significant (ΔAIC = 1, LLR χ^2 (1) = 1.023, p = .31). Numerically, when the relational lures were presented before their corresponding intact pairs, there were more false alarms (M = 0.21, SD = 0.41) compared to non-relational lures (M = 0.15, SD = 0.36). This was not the case however when they were presented after their corresponding intact pairs (both M's = 0.18, SD's = 0.38).

Discussion

This experiment presents the first empirical demonstration that specific relations in LTM can influence recognition judgments of lures in an associative recognition task. Specifically, we showed that people take longer to reject lures (e.g., [TABLE CLOTH]) that were relationally similar (*is covered by*) to word pairs that they had studied (e.g., [FLOOR CARPET]). We will refer to this result as *the relational luring effect*. Additional analyses demonstrated that the effect is not due to priming in short-term memory, because the effect occurred only when the lure was tested before its corresponding intact pair. This result suggests that relations may have abstract independent representations in LTM that can be retrieved by different instances of those relations. We posit that implicit semantic relations between words in pairs are encoded during the initial study phase, and that they serve as a spurious source of familiarity during the recognition of relational lures.

While our prediction about the effect of spurious relational information on RTs was supported, contrary to our expectations, we did not find a significant increase in false alarms. It could be argued that when it comes to memory and analogy, it is more important how accurate people are in their recall and their reasoning, rather than how fast they can retrieve the necessary information. For that reason, demonstrating the relational luring effect on false alarms would constitute a much stronger support for our hypothesis. It is possible that our experiment was underpowered, which prevented us from detecting an effect on accuracy. Another possibility is that the task was relatively easy and that due to the short list length, people were able to recognize lures despite the increased familiarity. Since participants studied only 21 word pairs before being tested on them, when they were presented with a recombined pair during test, they could have recalled that one of the words was originally studied in a different pair, and consequently use this recall-to-reject strategy in order to facilitate correct responses. Alternatively, having seen only a single instance of a relation might not induce enough familiarity in order to influence the type of the response, despite slowing down correct rejections.

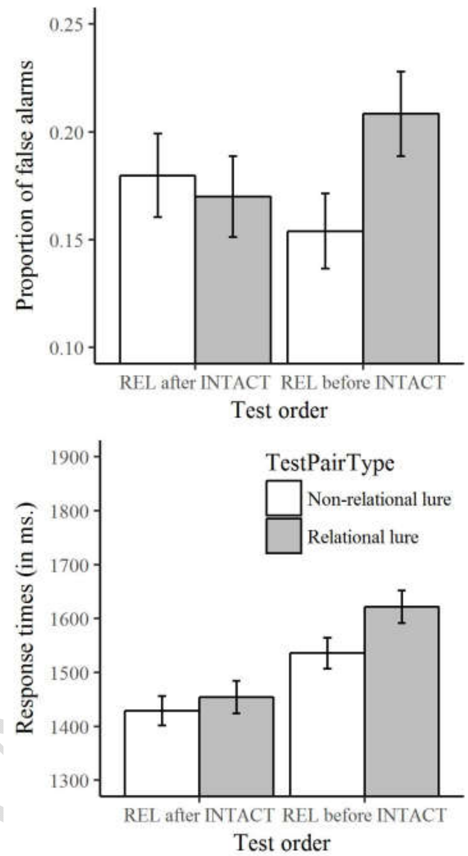

Figure 2. Proportion of false alarms (top panel) for relational and non-relational lures, and response times (bottom panel) in Experiment 1 for each type of pair, depending on whether the relational lure pair was tested after (REL after INTACT) or before (REL before INTACT) its corresponding intact pair. Error-bars represent ±1 SE.

Experiment 2

To overcome the limitations of Experiment 1 and to provide further support for the relational luring hypothesis, we performed a continuous pair recognition experiment with an expanded set of materials, where participants saw multiple exemplars of each relation. In continuous recognition tasks, study and test items are not split into separate blocks, but are instead intermixed throughout a single continuous sequence of trials. In the version of the task used in Experiment 2, new, recombined, and intact pairs were continuously introduced, and on each trial the participant had to respond whether the pair is "new", "recombined" or "old". During the beginning

 Journal of Experimental Psychology: General. DOI: 10.1037/xge0000305

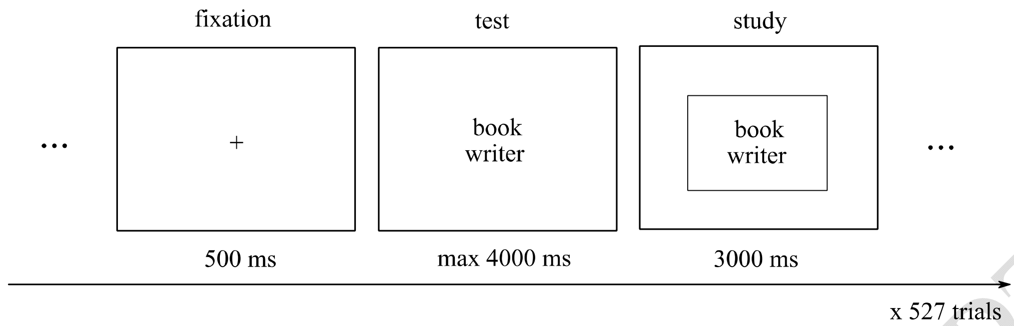

Figure 3. Example trial in Exp 2. The test screen remained until response or 4000 ms. elapsed.

of the experiment most pairs were new, but as the experiment progressed, the recombined and intact pairs appeared more frequently. This task was significantly longer and more difficult than the split study-test procedure in Experiment 1. More importantly, however, it allowed us to study how false alarms and RTs differ as a function of how many different exemplars of each relation had already been seen. In Experiment 1, there were only two types of recombined lures: either relational (1 previously seen exemplar of the relation) or non-relational (0 previously seen exemplars of the relation). In Experiment 2, we manipulated the strength of the relational lures (and targets) by presenting 1, 2, 3 or 4 different exemplars of the relation *prior* to the recognition test of the lure or the target.

If the relational luring hypothesis is correct, then every time a participant sees a different exemplar of a relation (e.g. the exemplars BOOK WRITER, BLUEPRINT ARCHITECT, PAINTING ARTIST, SONG COMPOSER, etc, for the relation "*is created by*"), the activation strength of that relation in LTM should increase. Since recognition memory is strongly influenced by the activation strength of items in LTM (Reder et al., 2000), the activated relation might cause spurious familiarity when responding to other exemplars. If that is the case, then we expect that as the number of previously seen exemplars of the relation increase, the RTs should increase for recombined pairs (H1), but not for the intact pairs (H2), and that the proportion of "old" responses should increase for both recombined pairs (i.e., increased false alarms; H3) and intact pairs (i.e., increased hits; H4). RTs for recombined pairs should increase because the spurious familiarity is in conflict with identifying the pair as "recombined", not as "old". In contrast, for intact pairs RTs might remain constant or they might even decrease because familiarity supports the correct "old" response. While we can predict that RTs should not increase for intact pairs, whether they remain constant or whether they decrease would depend on the degree to which responses are based on familiarity vs recollection (we will return to this point in the General Discussion). Finally, "old" responses should increase with the number of previous exemplars of the relation for both recombined and intact pairs, due to the increased familiarity of the relation. Support for *all four* aspects of this prediction would constitute a strong test of the relational luring hypothesis.

Method

Participants. Twenty-one undergraduate students (4 males) at New Bulgarian University participated for partial fulfillment of course credit and for the chance to win a gift card worth approximately \$13, \$10 and \$8 (for the three people with highest accuracy). All were native Bulgarian speakers, whose age ranged from 18 to 43 years (M = 23.3, S D = 6). The study was approved by the Department of Psychology Ethics Committee.

Procedure. We used a continuous pair recognition task where new, recombined, and intact trials were intermixed into a single continuous procedure in which there were no separate study and test sessions. Each trial began with a fixation cross for 500 ms. The fixation was followed by two words that appeared one above the other in the middle of the screen. Participants were instructed to respond immediately after the presentation of the word pair whether they thought they had previously seen the two words in the same pair ("old" response), whether the words had been seen in separate pairs ("recombined" response), or whether the words were never seen before in the experiment ("new" response). Participants had up to 4 seconds to respond, and if they did not press the corresponding button within that time, a message was displayed reminding them that a response is expected. When a response was made, or after they dismissed the reminder in case of no response, a border box appeared around the word pair and it remained on the screen for another 3 seconds, which gave the participant a chance to study it. This additional time for study was implemented so that RTs will reflect only memory retrieval plus decision speed, and not a mixture of that and studying the word pairs. In this way, participants were able to respond as quickly as they could, without worrying that they will have to remember the pair for later. Participants were explicitly instructed to respond before studying the pair, and to be maximally accurate, while responding within 4 seconds. After the 3 seconds for the study elapsed, the next trial began (the full procedure is illustrated in Figure 3).

The main task was preceded by two training phases. The first training phase introduced the mapping between the three response options and the three corresponding buttons, and was implemented in order to reduce interference during the

Table 3 Example partial trial order for a set of exemplars for a single relation in Experiment 2. Note the trial numbers – we have omitted intermediate trials for the illustration, but other pairs were presented in between the trials below. Nexemp = "number of previously seen exemplars of the relation". The "Lure?" column shows the type of lure in the terms of Experiment 1 for comparison.

Trial #	Pair	Trial type	Relation	Nexemp	Lure?
12	ring bank	new	-	-	-
14	cashier conductor	new	-	-	-
73	nurse gram	new	-	-	-
81	screwdriver hospital	new	-	-	-
95	vendor cold	new	-	-	-
109	weight shop	new	-	-	-
154	nurse hospital	recombined	works in	0	Non-relational
185	**nurse hospital**	**intact / old**	**works in**	**0**	**-**
217	cashier bank	recombined	works in	1	Relational
225	mechanic clothes	new	-	-	-
230	couch workshop	new	-	-	-
254	**cashier bank**	**intact / old**	**works in**	**1**	**-**
260	waiter night	new	-	-	-
266	birth restaurant	new	-	-	-
341	vendor shop	recombined	works in	2	Relational
374	**vendor shop**	**intact / old**	**works in**	**2**	**-**
421	waiter restaurant	recombined	works in	3	Relational
473	**waiter restaurant**	**intact / old**	**works in**	**3**	**-**
496	mechanic workshop	recombined	works in	4	Relational
518	**mechanic workshop**	**intact / old**	**works in**	**4**	**-**

main task. A word reflecting the expected response ("new", "recombined", "old") appeared in the middle of the screen and the participant had to press the corresponding button. All response options were tested 10 times and were presented in a random order. The second training phase introduced the continuous recognition task. It had 20 trials and the only differences between it and the main task was that accuracy feedback was given after each response. In the main task, feedback about the participant's average accuracy was given after every 50 trials. There were a total of 527 trials in the main task and the whole session lasted approximately an hour.

Materials. The final set of stimuli included 165 unique word pairs and each word pair was an instance of 1 of 35 unique relations. Each relation was represented by four (10 cases) or five (25 cases) exemplars[1]. For example, the relation "*X works in Y*" was represented by five word pair exemplars – NURSE HOSPITAL, WAITER RESTAURANT, VENDOR SHOP, CASHIER BANK and MECHANIC WORKSHOP (the full list of stimuli is presented in Appendix A). It was important that each word pair is a strong and typical exemplar of its relation, so the stimuli were pretested via a norming study with a separate sample of participants (the full method of the pretesting is described in Appendix B). In summary, in the pretest we presented participants with 2 exemplars for each of 58 relations, which were already determined to be dominant exemplars of their relations in a previous study (Popov & Hristova, 2015), and we asked them to generate for each set 3 novel exemplars that have the same relation. From their responses we selected the 2 or 3 most often produced exemplars, under the constraint that across all exemplars all individual words must be unique.

In the main experiment, the final set of 165 word pairs were never shown as "new" trials. Instead, the new trials were constructed by randomly recombining the individual words from the 165 different pretested pairs. Then, the pretested pairs were presented twice. During their first presentation they were "recombined", because the individual words had been seen in separate new pairs. During their second presentation they were "intact", because the individual words had already been seen together in the previous presentation. These "new", "recombined" and "intact" pairs were intermixed together throughout the experiment. This design, in which "intact" pairs were always repetitions of recombined pairs, was chosen to maximize power, while keeping the experiment as short as possible – in that way all exemplars of each relation can be used both as recombined and as intact pairs. In addition, there was another set of unrelated 32 words pairs that served as fillers in "new" trials in the last third of the experiment. Hence, in total, there were 197 new, 165 recombined and 165 word intact pairs that were intermixed together in the trial sequence. Participants were instructed that they should respond "old" if a pair is repeated/intact, even if it was recombined the first time they saw it.

An example of the trial order for one set of exemplars of a single relation is presented in Table 3. For example, some time after seeing the new pairs RING BANK, CASHIER CONDUCTOR, NURSE GRAM and SCREWDRIVER HOSPITAL ("new" trials), the words NURSE and HOSPITAL were presented together and a "recombined" response was expected. Subsequently, the word pair NURSE HOSPITAL was presented again ("intact" trial), and an "old" response was expected. Since this was the first exemplar of the relation "*works in*" (0 previously seen exemplars), this was equivalent to a "non-relational lure" in Experiment 1. Later on the words CASHIER and BANK were recombined and presented together. In this case, the word pair CASHIER BANK was the second exemplar of the relation and was equivalent to a "relational lure" in Experiment 1. Similarly, the subsequent exemplars were also relational lures, with each one supposedly being a stronger lure, due to the multiple repetitions of the relation.

By necessity most of the pairs in the beginning of the experiment were new, while recombined and intact pairs started to appear for the first time on average after 26 trials (ranging from 23 to 31) of new pairs. The order of the trials was randomized for each subject. To ensure that decisions to intact and recombined pairs are based on LTM rather than STM, the randomization was done via an algorithm that satisfied following constrains: 1) each recombined pair is presented at least 20 trials after the presentation of the new pairs that contain the words in that recombined pair (M = 179, SD = 113, range = 21 to 500); 2) each intact pair is presented at least 20 trials after its previous occurrence (M =35, SD = 22, range = 21 to 219); 3) there is a trial lag of at least 20 trials between presentations of different exemplars of the same relation to prevent grouping strategies; 4) no new exemplars of a relation appear between a specific recombined exemplar and its repetition as an intact pair; and 5) there are no more than 4 consecutive intact or recombined trial types (M = 1.5, SD = 0.8).

Design. We used a within-subject design with 2 main factors – type of trial ("new" vs "intact" vs "recombined"), and number of previously seen different exemplars of the relation (0 vs 1 vs 2 vs 3 vs 4). It is important to stress that we are talking about the number of previously seen *different* exemplars of the relation in the currently tested pair, not the number of repetitions of that specific pair. In the terms of Experiment 1, the zero previous exemplars condition corresponds to a "non-relational lure" when the trial is recombined, and the remaining levels correspond to relational lures of increasing strength. Across participants each word pair was used equally often in each condition for the number of previous exemplars.

Results and Discussion

Response times. The RTs were analyzed via linear mixed-effects regression models (Baayen, Davidson, & Bates, 2008). The random effects were determined through restricted likelihood ratio tests and the final model consisted of varying intercepts for subjects, individual word pairs and groups of relations (i.e., different subjects, items and relation groups vary in their overall RT estimates), as well as varying slopes by subject and by relation group for the effect of type of trial and number of previous exemplars (i.e., the model accounts for how much the relational luring effect varies across subjects and relations). None of the fixed effect estimates were significantly affected by varying the random structure. There were no differences in the conclusions between analyzing raw and log-transformed RTs, and we report the raw analyses for ease of interpretation. Only accurate

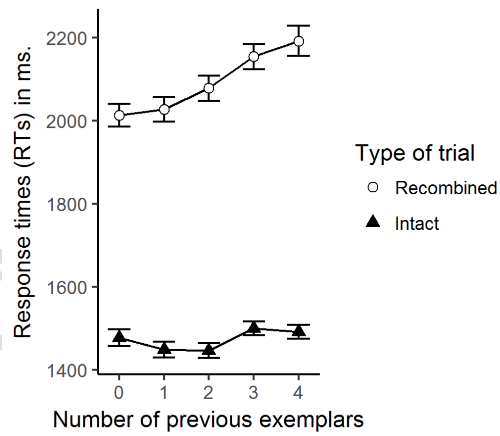

Figure 4. Response times in Experiment 2 as a function of type of trial and how many exemplars of the relation participants have
al r the effect

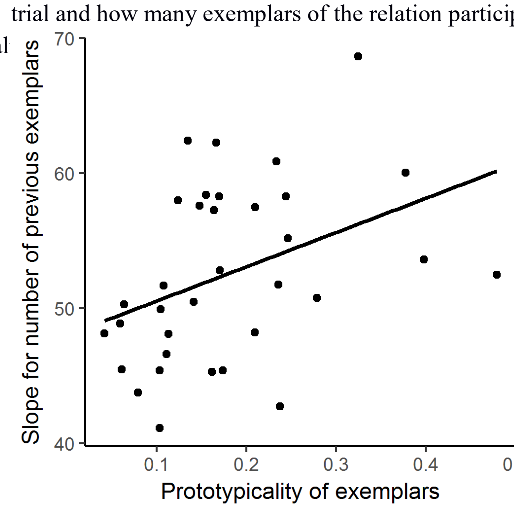

Figure 5. The effect of the "number of previous exemplars" in Experiment 2 increases when the exemplars are more prototypical of their relation. Each point is one relation group of exemplars, and the y axis is the random slope of number of previous exemplars for each relation group estimated in the mixed-effect regression model. On the y axis is the mean proportion of people in the pretesting study that generate each exemplar of the relation.

Table 4 Mixed-effects regression coefficients and model comparisons for RTs in Experiment 2. The estimates of the model are in miliseconds. *Recombined trials were the reference level and thus the Trial type (intact) beta value corresponds to the difference on intact trials relative to the intercept for recombined trials. ^Nexemp = “Number of previously seen exemplars of the relation”. Nexemp and trial number were mean-centered.

Parameter	**Estimate**	**95% CI**
Fixed effects		
Intercept *	2216	[2019; 2240]
Trial type (intact) *	-650	[-753; -545]
Trial number	-0.7	[-1.1; -0.4]
Nexemp^	52	[13; 90]
Trial type (intact) x Trial number	-0.5	[-0.9; -0.1]
Trial type (intact) x Nexemp	-50	[-92; -7]
Random effects variance		
Word paid (Intercept)	5883	
Relation (Intercept)	2898	
Relation (Nexemp)	190	
Subject (Intercept)	65847	
Subject (Nexemp)	1325	
Subject (Intact trial type)	50429	
Residual Variance	191334	

Parameter	**ΔAIC**	**χ^2 (df)**	***p***
Trial number	-70	72.03 (1)	< *.001*
Trial type	-44	46.56 (1)	< *.001*
Trial type × Trial number	-75	76.66 (1)	< *.001*
Nexemp	0	1.97 (1)	*.159*
Trial type × Nexemp	-3	5.21 (1)	*.022*

responses (71%) were included in the RT analyses. Inferences about significance were made based on likelihood ratio tests and AIC comparisons of the regression model that compared the effect in question to an identical regression model that lacked only this effect. Since the number of previous exemplars was correlated with trial number, it was important to control for this in the model. Fixed effects were added step-wise to the regressions in the following order: Trial number (mean centered), Trial type, Trial number (mean centered) × Trial type, Number of previous exemplars (mean centered), Number of previous exemplars (mean centered) × Trial type. The coefficients and the statistical tests for the mixed-effects regression of RTs are presented in Table 4. Below we discuss the effects in turn.

There was a relational luring effect on RTs for recombined pairs, but not for intact pairs (see Table 4 and Figure 4). People needed more time to correctly recognize recombined pairs as “recombined” for relational lures (1 to 4 previous exemplars of the relation), compared to non-relational lures (0 previous exemplars of the relation). Importantly, the relational luring effect increased with each novel exemplar of the relation – every previously-seen exemplar of the relation increased response times to the current *recombined* pair by 52±20 ms. This result is directly comparable to the 61 ms. difference between the relational and non-relational lures in Experiment 1. In addition to replicating the relational luring effect, Experiment 2 extends the findings by showing that the effect increases with the strength of the relation in the lures, as defined by the number of exemplars that have already activated it.

Importantly, the relational luring effect was not due to a general slow-down over the course of the experiment, because there was an opposite practice effect in that RTs decreased with every trial, and that speed-up was greater for intact pairs (1.2±0.2 ms) than for recombined pairs (0.7±0.18 ms). Thus, the slow-down effect of number of different previously-seen exemplars of the relation was obtained despite a general speed-up over trials.

We can also ask whether the relational luring effect varies in strength depending on *how typical on average* the exemplars are for each relation. Exemplars that are more typical might have a greater chance to activate their relation, or might activate it more strongly, which in turn would be more likely to cause spurious familiarity during processing subsequent exemplars. To test this idea, one would need to compare 1) a measure for the average typicality of the exemplars for each relation, with 2) a measure of the strength of the relational luring effect for each relation set. Since our stimuli were pretested, we can define typicality as the mean proportion of participants that generated each exemplar, and then take the average of that value for a measure for the whole set of exemplars of the relation. While generation frequencies are not a direct measure of typicality per se, they have been widely used as a proxy for typicality due to the high correlation between them (Mervis, Catlin, & Rosch, 1976; Rosch & Mervis, 1975). When it comes to the strength of the effect, the mixed-effects regression analysis fitted a separate estimate of the relational luring effect for each relation set (the random slope estimates for the "number of previous exemplars" effect). As can be seen from Figure 5, the strength of the relational luring effect on RTs increased when the exemplars were more typical of the relation, $r(34) = 0.39$, p = .02 (one data point was removed because a residual plot reveal it was an outlier (standardized residual > 3). This result further characterizes the relational luring effect, by showing that variability in its strength is directly related to variability in the underlying relational representation that causes it.

Responses. The type of response was a 3-level categorical variable ("new", "old" or "recombined"), which we analyzed with a multinomial mixed modeling framework that was estimated by Markov chain Monte Carlo methods with the R *MCMCglmm* package (Hadfield, 2010). The multinomial logistic regression model is an extension of binary logistic regression for dependent variables with $k > 2$ levels. Its exponentiated estimates, known as *relative risk ratios* (RR), are the probabilities of each of *k-1* levels *relative to a reference (baseline) level*. The RR has a similar interpretation to odds ratios (OR)[ii] in binary logistic regression – in which the *relative risk (probability)* of making a choice k_i compared to a reference choice k_{ref} is $\frac{P(k_i)}{P(k_{ref})}$. For categorical predictors, the multinomial model estimates a separate log RR for each level of the predictor and for each response option relative to the baseline. However, the estimate for continuous predictors, such as our "number of previous exemplars of the relation", represents a multiplicative change in the RR for a unit change in the predictor.

In the current experiment we set the "recombined" response as the reference level (baseline), and the model estimated posterior log RR of an "old" or a "new" response relative to a "recombined" response. The fixed effects included the type of trial, the trial number separately for each trial type, the number of previous exemplars of the relation,

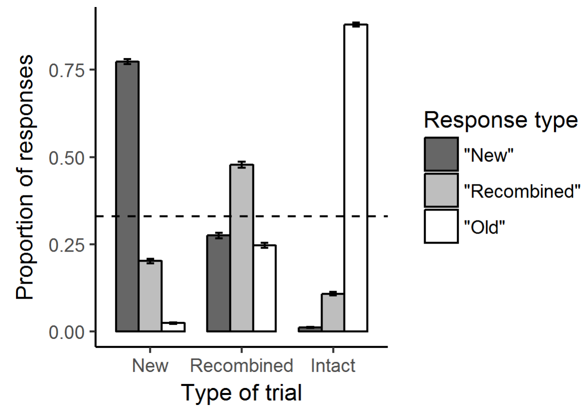

Figure 6. Proportion of responses in Experiment 2 for each type of response for each type of pair. The horizontal line represents chance performance.

and their interaction with type of trial. The random effects in the final model included varying intercepts for each trial type by subject, as well as by-subject varying slopes for the number of previous exemplars in each trial type. The random effect structure was determined by comparing the DIC values of alternative models that converged successfully. Due to high autocorrelation and slow convergence of the random effects and the DIC value, all models were run with 3 chains for 1,000,000 steps with a burn-in of 10,000 steps and a thinning factor of 200. Inspection of convergence diagnostics (e.g., traceplots, $\hat{R}$'values, running means) suggested appropriate convergence for all parameters (e.g., all $\hat{R}$'s < 1.03). The priors for the fixed effects were uninformative normal distributions centered around 0 with $SD=10^8$, and the random effects had uninformative Cauchy priors with a scale of 25 (Gelman, 2006). Finally, conclusions about the parameter values were drawn based on whether 0 (or 1 for the exponentiated values) lied within the 95% *credible intervals* (CI) of the posterior distributions of each parameter (Kruschke, 2010).

The descriptive results are shown in Figures 6 through 9, and the relative risk ratios that were estimated by the multinomial mixed-effects regression for each effect and their confidence intervals are presented in Table 5. As can be seen from Figure 6, overall performance in the task was good despite its length.

In contrast to Experiment 1, we found evidence that relational information influences not only RTs, but responses as well (see Figure 7). Specifically, false alarms for recombined pairs were greater for relational lures (1 to 4 previous exemplars of the relation) compared to non-relational lures (0 previous exemplars of the relation). Importantly, as with RTs, the relational luring effect increased with each novel exemplar of the relation. For recombined pairs, the relative risk of responding "old" rather than "recombined" increased by 1.26 (95CI: 1.03-1.57) times for every different exemplar of the relation that participants had seen (the relative risk of making "new" vs "recombined" responses did not change significantly). Hence, the main

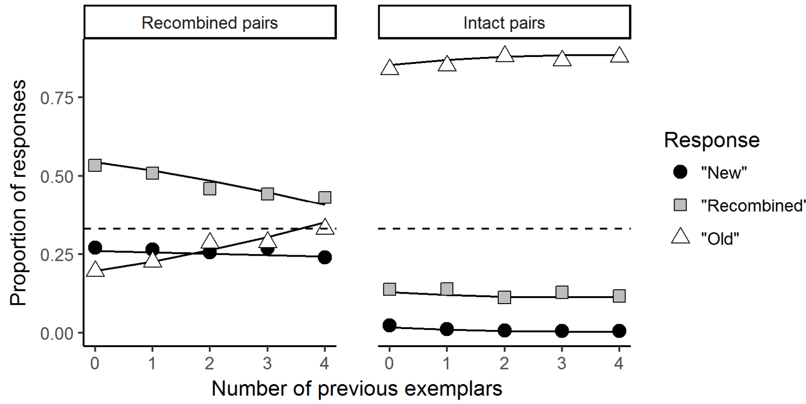

Figure 7. Proportion of each type of response for recombined and intact pairs depending on how many previous exemplars of the relation had been seen, while controlling for the effect of trial number (set to a constant). Solid lines represent the fit the of the multinomial regression, and the dashed line represents chance performance.

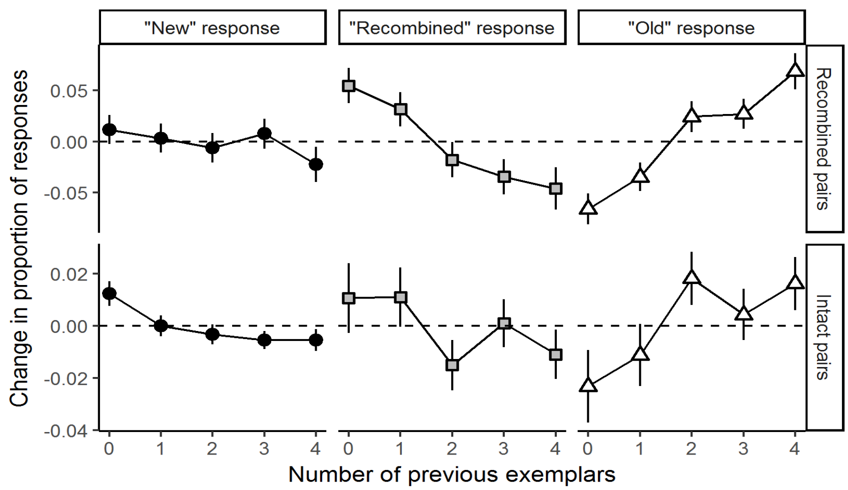

Figure 8. Change in proportion of each type of response in Experiment 2 for recombined and intact pairs depending on how many previous exemplars of the relation had been seen, while controlling for the effect of trial number (set to a constant). This is the same data as presented in Figure 7, but the mean proportion of each responses for each trial type was subtracted from each point to "zoom in" on the effect. Note the difference in the y-axis scales for recombined and intact pairs. The dashed line represents the mean in each panel. Errorbars show ±1SE.

predictions of the relational luring hypothesis, that false alarms to recombined pairs should increase when people have already seen the relation in different exemplars, and that the effect should increase with increasing the number of exemplars, were supported.

A story consistent with the relational luring hypothesis emerged for intact trials as well, although the effect was weaker. Since small changes in response proportion were hard to see on the raw proportion scale in Figure 7, we "zoomed in" on the effect by replotting the data in Figure 8. Specifically, we subtracted the mean proportion of responses of each type separately for each trial type. The resulting scale shows how much the proportion of responses changes with each seen exemplar, relative to the mean of the specific response and pair condition. For intact pairs, the likelihood of responding "old" relative to "new"[iii] increased by 2.25 (95CI: 1.25-3.71) as more exemplars of the relation had been previously seen. Even though for intact pairs, this effect is not relational *luring* per se, because it leads to better performance, it is consistent with it. Activating the relation through multiple exemplars makes it stronger and just as that leads to more "old" responses for recombined pairs, it does the same for intact pairs. One possible reason that the effect is much smaller for old pairs, is that participants might be using recollection rather than familiarity for responding to the majority of old pairs, and then switch to familiarity only when recollection fails them. We will elaborate on this point in the gene ral discussion.

Finally, the increased likelihood of responding "old" with each novel exemplar was not due to a confound with trial number. As with the RTs, for recombined pairs, the effect of number of exemplars was observed despite the opposing

Table 5 Posterior means and 95% credible intervals for the relative risk ratios (RR) from the bayesian mixed-effect multinomial logistic model of response types in Experiment 2. “Recombined” responses were the reference category. See the *data analysis* section for more details. Nexemp = “number of previously seen exemplars of the relation. Bolded are RR whose CI does not include 1.

Parameter	"New" vs "Recombined" responses RR (95% CI)	"Old" vs "Recombined" responses RR (95% CI)
Recombined pairs		
Constant	**0.204 (0.090 - 0.467)**	**0.423 (0.215 - 0.837)**
Trial number	**1.002 (1.000 - 1.004)**	**0.998 (0.997 - 1.000)**
Nexemp*	0.958 (0.765 - 1.196)	**1.268 (1.026 - 1.566)**
Intact pairs		
Constant	**0.014 (0.003 - 0.061)**	**3.224 (1.477 - 6.953)**
Trial number	**1.008 (1.003 - 1.014)**	**1.005 (1.003 - 1.008)**
Nexemp*	**0.491 (0.283 - 0.854)**	1.064 (0.826 - 1.360)

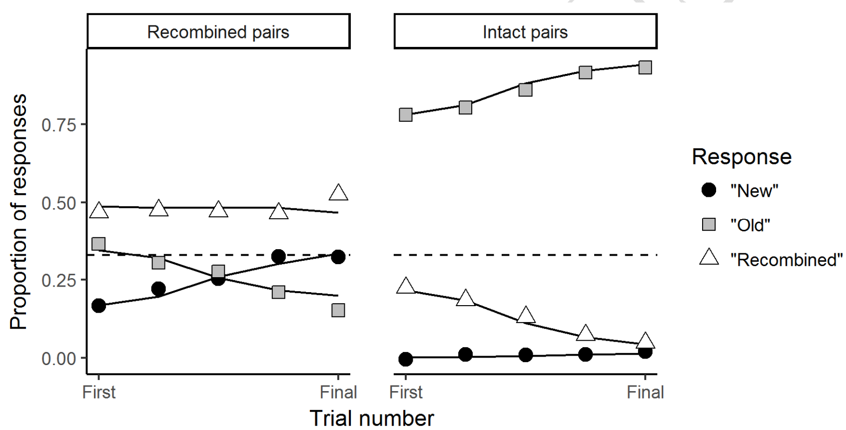

Figure 9. Proportion of responses in Experiment 2 for each type of response for each type of pair as a function of the trial number / progress through the experiment. Solid lines are the fits of the multinomial regression model. The horizontal line shows chance level.

effect of trial number. Figure 9 shows how responses change for intact and recombined pairs throughout the experiment. In general, as the experiment progressed, participants became better at recognizing intact pairs as “old”, and “recombined” responses to intact pairs decreased with time. At the beginning of the experiment participants were 3.2 times more likely to give an “old” rather than a “recombined” response to intact pairs, but with each trial this likelihood increased by 0.5%iv. New responses were at floor for intact pairs throughout the experiment. For recombined pairs the results are quite different. In the beginning of the experiment, participants were 4.9 times more likely to respond “recombined” rather than “new”, and 2.4 times more likely to respond “recombined” compared to “old”. However, their accuracy did not improve over trials, only the error types changed. Specifically, with each trial the relative risk of false alarms decreased by 0.2%, while the relative risk for “new” errors increased by 0.2%, leaving their overall performance unchanged. Thus, despite that “old” responses to recombined pairs decreased as the experiment progressed, they increased with each previously-seen exemplar of the relation.

Summary. As in Experiment 1, participants took more time to reject recombined lures that were relationally similar to previously seen pairs and the effect increased linearly with the number of previously seen exemplars of the relation (H1). In contrast, RTs for intact pairs remained constant (H2). A relational luring effect was observed in the response types as well on both intact and recombined pairs. As the number of previously seen exemplars of the relation increased, participants became more likely to respond “old” when a word pair was relationally similar to previously seen pairs, regardless of whether the pair itself was presented for the first time (recombined pairs; H3), or if it was repeated (intact pairs; H4). The effect on RTs or the responses was not due to a confound with trial number. Finally, the effect on RTs was stronger when the exemplars were more typical of the relation. Combined, these findings suggest that the relational luring effect is continuous in nature, rather than binary, and that its magnitude depends on the activation strength of the relational in LTM.

General Discussion

Based on theoretical considerations and previous empirical work, we proposed that semantic relations are represented abstractly in LTM and that these abstract representations can be retrieved by different instances of those relations. Experiment 1 supported this hypothesis by demonstrating a novel relational luring effect during associative recognition. Particularly, when participants were tested on recombined word pairs [TABLE CLOTH] that were relationally similar (*is covered by*) to some of the studied pairs [FLOOR CARPET], they needed more time to correctly reject the relationally similar lures and they had a non-significant tendency to make more false alarms. Since the relation encoded in the study pair influenced judgments on a word pair composed of different items, the relational luring effect suggests that the representations of the relations in LTM are abstract and independent from the entities that instantiate them. It further suggests that during recognition the initially encoded relation is retrieved and serves as a false source of familiarity, which leads to slower response times for correct rejections.

Even stronger support for the relational luring hypothesis came from Experiment 2, which presented multiple different exemplars of each relation before testing recognition memory for novel exemplars. Based on previous research, we know that stronger activation of single items in LTM leads to increased familiarity during recognition (Reder et al., 2000). We hypothesized that if relations also have abstract representations in LTM, then each different exemplar of a relation should activate the same representation. If that is the case, then seeing more than one exemplar of a relation should lead to the accumulation of activation and to increases in familiarity of the relation. We expected this increased familiarity to result in increasing the magnitude of the relational luring effect proportionally to the number of previously seen exemplars of the relation.

In Experiment 2 this was the case for both RTs and for false alarms. Specifically, every time participants saw a different exemplar of a relation (e.g. NURSE HOSPITAL, WAITER RESTAURANT, VENDOR SHOP, CASHIER BANK and MECHANIC WORKSHOP for the relation "*X works in Y"*), this caused them to be on average 52 ms slower in correctly identifying the subsequent exemplar as recombined. Importantly, the relational luring effect accumulated – for instance, seeing 4 exemplars caused participants to be one average 208 ms slower on the fifth exemplar). Similarly, each new exemplar made participants more likely to false alarm on the subsequent one. After seeing 4 different exemplars of a relation participants were almost equally likely to say that they had seen the 5th exemplar before, as they were to correctly identified it as recombined (a 2.6 times increase in false alarms vs correct rejections). Both of these effects are a clear indication that each exemplar accessed the same representation of the relation in LTM.

Furthermore, we found that the relational luring effect was modulated by how typical the exemplars were of their relation. Specifically, relations that were represented by more typical exemplars, lead to a greater increase in RTs for correct rejections of recombined lures. While we did not predict this effect, it is largely consistent with the relational luring hypothesis. More typical exemplars (e.g. NURSE HOSPITAL) of a relation (*works in*) might be more likely to activate it than less typical exemplars (such as STEWARDESS AIRPLANE), and as we have already established, activation strength of the relation is proportional to the effect. It is worth noting that we did not measure typicality per se, but rather how frequently each exemplar was generated by a separate sample of participants. It unlikely that this difference changes the interpretation of the results since generation frequency has often been used as a proxy for typicality due to their high correlation (Mervis, Catlin, & Rosch, 1976). Importantly, the relational typicality result uncovered here is in direct contrast to previous studies that have failed to find any consistent differences in the availability or ease of processing of different relations (Gagné et al., 2005; Shoben, 1991).

Why would less typical exemplars of a relation be less likely to activate it? While we are not aware of any research that deals specifically with the typicality of *relational* exemplars, it has been well established that the prototypicality of category exemplars (such as *apple* for the category *fruit*) plays a huge role in categorization speed, age of acquisition, etc (Rosch, 1999; Rosch, Simpson, & Miller, 1976). Furthermore, researchers have argued that prototypicality and generation frequency are indices of the accessibility of concepts in LTM (Janczura & Nelson, 1999). Many models of LTM account for these effects by varying the strength of the links between the category and the exemplar representations. Similarly, activation of the relation node in LTM might be a function of the strength of the association between it and its exemplars.

Our final prediction for Experiment 2 concerned the effect of increasing the number of exemplars on previously seen intact pairs. In contrast to recombined pairs, we did not expect to find an effect of relational similarity on the RTs for intact pairs. Intact pairs required an "old" response, and as such increasing the familiarity of the relation should not create a conflict that needs to be overcome during the decision-making process. However, the increased familiarity lead to a greater probability of correctly identifying the intact pairs as old. On the surface, this effect is not strictly relational *luring*, because it leads to better performance. However, we will continue to refer to it as such even for intact pairs, because the underlying cause is the same. As with the recombined pairs, the increase in "old" responses is not due to better encoding or retrieval of intact pairs, but it is instead due to a source of information external to the specific item being tested, namely, the activation of the relation from other exemplars. This activation produces a spurious source of familiarity, which for intact pairs happens to coincide with expected response.

Familiarity vs Recollection

It is notable, however, that the relational luring effect for the responses was much smaller for intact pairs compared to recombined pairs. Why would that be the case, given that the activation of the relation caused by the previous exemplars should be equal in both cases? One simple explanation comes from considering the two main decision processes in dual-process models of recognition memory, familiarity and recollection (for reviews, see Diana, Reder, Arndt, & Park, 2006 and Yonelinas, 2002). Dual-process models suggest that familiarity decisions are based on assessing the generic strength of items in LTM. Alternatively, recollection involves the retrieval of specific mnemonic traces and contextual associations created during the study episode. Within this framework, the correct identification of recombined pairs depends mostly on familiarity, because they have not been seen before during the experiment and as such no episodic details can be recollected. In contrast, the correct recognition of previously seen pairs can depend both on familiarity and recognition. Furthermore, some models such as SAC (Reder et al., 2000) posit that familiarity is used for recognition, only if an initial attempt at recollection fails.

The idea that the dependence on familiarity is much greater for recombined pairs naturally explains why the relational luring effect is also greater for them. Our original hypothesis was that the relational luring effect would be the result of spurious familiarity caused by being exposed to other exemplars of the relation. Since, in contrast to recombined pairs, the recognition of intact pairs is only partially influenced by familiarity, any spurious familiarity from relational luring is expected to have a smaller effect. Thus, this difference in magnitude in the relational luring effect provides further evidence that its cause is indeed rooted in extraneous activation of a common representation of the relation by other exemplars.

Implications for models of LTM

The main theoretical contribution of this study pertains to its implications for models of memory and analogical reasoning, and specifically for the ways these models deal with the representation and retrieval of relational information. This study constitutes the first empirical demonstration of the relational luring effect, namely that specific semantic relations in LTM that are activated by some exemplars of a relation can influence recognition judgments of different exemplars as lures and targets in an associative recognition task. Aside from uncovering that a relational luring effect exists in recognition memory, this study further characterized the various ways in which it interacts with relational typicality, relational strength and type of trial. These effects constitute a rich new empirical constraint that needs to be accounted for in current and future models of LTM.

The Introduction identified three ways in which existing memory models deal with relations. Some network models (Collins & Loftus, 1975; Collins & Quillian, 1969) and some feature-based models (Smith et al., 1974) represent relations as being stored locally within the representation of entities, and in these models, relations do not have independent abstract representations. These types of models are incompatible with the relational luring effect because they cannot explain how one instance of a relation might influence judgments on another instance of that relation, if they are represented separately from one another.

The second type of models, primarily posed as theories of episodic memory, do not make any claims about how relations might be represented and retrieved (TCM: Howard & Kahana, 2002; CMRM: Polyn, Norman, & Kahana, 2009; SAM: Raaijmakers & Shiffrin, 1981; SAC: Reder et al., 2000; REM: Shiffrin & Steyvers, 1997; and others). Our results are especially relevant for episodic memory models such as these, because one of their explicit goals is to explain how people make associative recognition judgments, which is the task that we employed.

None of these models would have predicted the relational luring effect, because currently they do not address how preexisting semantic relations are represented in LTM. Some of them, however, could be modified to account for the RLE. Our own theoretical position is similar to SAC (Reder et al., 2000), which currently assumes that during the study of paired-associates one semantic node represents each item, and the novel *episodic* association between them is represented by a novel episode node that connects them to one another. Associative recognition judgments in SAC are made either by evaluating the strength of the item nodes, which results in familiarity-based responses, or by retrieving the episode node, which results in recollection-based responses. In SAC, however, all of the sematic information about an item is contained within its own node. Where our proposal differs from SAC is in that we believe semantic relations between items also have their own independent nodes, which are connected to all of their instances. Then when related word pairs are studied, the node for the semantic relation is encoded and linked to the current episode node alongside the item nodes. Finally, during the associative recognition of lures that share the same relation, this relational semantic node is retrieved and it causes spurious familiarity. Clearly, the relational luring effect will be an important novel benchmark for models of episodic memory.

Finally, even though they are not concerned with associative recognition tasks, proposition-based models of semantic memory, such as ACT-R (Anderson & Lebiere, 1998), and analogy, such as SME (Falkenhainer et al., 1989) and MAC/FAC (Forbus et al., 1995), as well as hybrid symbolic-connectionist models like LISA (Hummel & Holyoak, 1997), DORA (Doumas et al., 2008), and AMBR (Kokinov & Petrov, 2001), or even some modern semantic network models (Rogers & McClelland, 2004), are potentially more consistent with our results because they assume that relations do have independent abstract representations in memory. Despite being potentially consistent with our data, as of yet these models have not explicitly demonstrated an ability to spontaneously prime relations, or that relations can influence recognition judgments of lures. Ad hoc consistency is not a particularly strong criterion for model selection, and

 Journal of Experimental Psychology: General. DOI: 10.1037/xge0000305

the relational luring effect we identified presents a clear motivation for future theoretical work.

Implications for other empirical studies of associative recognition

The theoretical framework presented in this paper might offer an alternative explanation to a recent series of associative recognition studies on ‘unitization’ (Parks & Yonelinas, 2014). Researchers have shown that compound word pairs (Zheng et al., 2015) or word pairs that are associatively related (Kriukova, Bridger, & Mecklinger, 2013; Rhodes & Donaldson, 2007) produce a greater early (300-500ms) mid-frontal ERP old/new effect compared to unrelated word pairs during associative recognition. This early mid-frontal ERP old/new effect supposedly reflects familiarity based-responses (Rugg & Curran, 2007). Additionally, fMRI studies (Ford, Verfaellie, & Giovanello, 2010; Haskins, Yonelinas, Quamme, & Ranganath, 2008) have revealed that while the recognition of unrelated word pairs depends on the left hippocampus, the recognition of compound pairs is predicted by activity in the left perirhinal cortex, an area involved in familiarity responses (Mayes, Montaldi, & Migo, 2007). Haskins et al.’s (2008) study is particularly notable since the compound pairs were unrelated in general (“slope bread”), but the experimental design established a novel relation between them by providing definitions such as “a pastry eaten by mountain climbers.” Thus both EEG and fMRI studies suggest that the recognition of compound word pairs involves familiarity-based responses.

Because familiarity is thought to underlie only item recognition (Parks & Yonelinas, 2014; Yonelinas, 2002), these results are usually attributed to a unitization of the compound pair into a single item. An alternative explanation is suggested by our results. Compound pairs (e.g. *airplane pilot)* or associatively related word pairs (e.g. *dancer – stage*) contain a specific semantic relation between the two items. If this relation is encoded during the initial study phase, as our data suggests, then this relation could contribute to the increased markers of familiarity during recognition. Our study provides clear evidence that this occurs during the false recognition or correct rejection of lures, and Experiment 2 showed that even the correct recognition of repeated pairs benefits from having been exposed to the relation in a different word pair. Importantly, the unitization hypothesis cannot account for our results, because our stimuli were in Bulgarian, and in contrast to English, noun-noun combinations are not grammatical phrases in Bulgarian, and *cannot be perceived as a single unit*. We suggest that results usually attributed to *unitization* might be better explained by the activation of the relation that holds between the two words, rather than by perceiving the two words as a single unit. The two processes might also interact, and to understand their relationship we would need to perform additional carefully-designed studies that compare recognition for unrelated and related words pairs that can and cannot be unitized.

Implications for research on constructive memory / false memory

A principal finding of the current work is the evidence that relational similarity, just like object-based similarity, can guide memory decisions and reconstruction. That possibility was largely neglected in constructive memory research just as in memory research in general (but see Feldman & Kokinov, 2009; Pavlova & Kokinov, 2014). It is generally accepted that the basic mechanisms for LTM reconstruction, apart from schema-based reconstruction, are based on semantic similarity (Feldman & Kokinov, 2009), as in the famous Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995; also see Montefinese et al., 2015). However, different episodes can also be similar with respect to their relational structure, which in turn can lead to memory intrusions of semantically dissimilar elements, which play similar roles. Both current experiments provide crucial evidence for this possibility by demonstrating false recognitions of word pairs that instantiate relation, which was already perceived although within a different word pair. Thus, relations can guide memory retrieval at least when object and relational based familiarity is alike.

Implications for the retrieval of analogues during analogical reasoning

Analogical reasoning is a difficult process and it has been suggested that one major reason for that difficulty is the so-called ‘retrieval gap’ (Holyoak, 2012). This gap refers to the difference in difficulty between mapping already-retrieved analogues and that of retrieving relevant relationally-similar structures from LTM. Once the structures have been retrieved it is a relatively straightforward process to map their corresponding elements, though it might be computationally expensive (Holyoak, 2012; Hristova, 2009; Popov & Hristova, 2014, 2015). However, retrieval of the analogues is complicated by the fact that any of countless relational structures in memory might be relevant for solving the task at hand (for a similar argument, see (Feldman & Kokinov, 2009).

The relative contribution of semantic and relational similarity to the retrieval of relevant analogues is controversial. Previous research has demonstrated that relational structures can be retrieved from memory more easily when there is considerable semantic similarity between items in those structures and the current stimuli (Keane, 1987), presumably because shared semantic features get activated in the base analogue and cause them to be retrieved (Forbus et al., 1995; Hummel & Holyoak, 1997). This principle of retrieval by domain similarity plays a major role in most analogy models (Reeves & Weisberg, 1994). However, even though traditionally retrieval by relational similarity has been considered to be more difficult and to play a lesser role compared to retrieval by semantic similarity (Gentner & Smith, 2012; Holyoak, 2012), some models such as AMBR (Kokinov & Petrov, 2001) and LISA (Hummel & Holyoak, 1997) do include mechanisms of relational retrieval.

Moreover, relational retrieval along with analogical mapping is a key mechanism in AMBR for spontaneous episode blending and memory reconstruction (Feldman & Kokinov, 2009). A specific prediction of the model was that people will tend to blend relationally similar episodes rather than or at least as much as superficially similar ones, which was indeed supported in subsequent behavioral experiments (Feldman & Kokinov, 2009; Pavlova & Kokinov, 2014).

From these and previous results the following picture emerges: in parallel to the activation of semantically similar elements, single relations within a relational structure might locally activate different instances of those relations, which are present in a variety of schemas in LTM. If one or several relational structures in LTM contain many relations or semantically similar elements that are activated locally by similar elements in the target, then the whole structure might be brought to mind for additional processing. This account is consistent with evidence that the induction of a schema improves relational retrieval (Dixon & Dohn, 2003; Gick & Holyoak, 1983), presumably because relations will be more clearly represented in a stable schema. It is also consistent with the fact that experts in a specific domain are more likely to retrieve analogues based on relational similarity (Novick, 1988), because their representations tend to be more organized around deep structural features (Cummins, 1993) and they might use relations as retrieval cues more efficiently.

One reason why a 'relational gap' exists might not be so much due to the difficulty of relational retrieval, but instead due to the fact that the representations of relations are often inconsistent across the base and target structures (Dunbar, 2001). During standard laboratory tasks of analogical reasoning, researchers expect that participants will encode the relational structure of the base and target situations correctly, but this assumption is rarely tested. Indeed, previous research has shown that when it is known that the relevant relations in the base and the target have been encoded correctly, participants are more likely to retrieve the base analogue (Catrambone & Holyoak, 1989; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983).). Recent research also shows that there are individual differences in how well relations are encoded and represented (DeWolf et al., 2017). For example, relational priming during mathematical problem solving occurs without the assistance of superficial similarity between the exemplars only for students that have high math ability (DeWolf et al., 2017). In summary, in order to claim that a genuine "retrieval gap" exists, researchers should demonstrate that the relational structure that they want participants to retrieve has actually been encoded in the first place.

Models of analogical reasoning do not show a retrieval gap because the relations in the base and target situations are coded by hand and thus the problem of how to represent relations in the input has been avoided all together. Some researchers believe that this is one of the fundamental shortcomings of current analogy models (Kokinov & French, 2003). We agree with this explanation for the "retrieval gap", because when relational similarity and typicality of exemplars are pretested and controlled for (e.g. Popov & Hristova, 2015, and the current study) relational retrieval occurs relatively easy and without intention. Finally, the current finding that relational luring is stronger for more typical exemplars of a relation presents clear evidence that the "retrieval gap" is indeed smaller when the word pairs are better exemplars of the relation.

Future directions

One aspect of our study is that we used an explicit associative recognition task, and participants might have tried intentionally to encode the relation between the items. This served well in our objective to demonstrate that such an effect exists, but an even stronger demonstration will involve an implicit memory task, where participants are not motivated to encode the relation between the entities. From our previous study (Popov & Hristova, 2015), we know that participants do encode relations between items in a word pair when it is irrelevant for the task, but we still do not know whether that encoding will be strong enough to influence associative recognition judgments after a significant delay. An implicit memory task might also make the task more difficult and increase false-alarms overall.

A related question of interest is how sleep might affect false memory for relational lures. Recent studies have found that false memory in the Deese-Roediger-McDermott paradigm is enhanced after a full-night sleep (Diekelmann, Born, & Wagner, 2010; Payne et al., 2009), presumably because sleep promotes integration of recently acquired memories into pre-existing long-term networks (Diekelmann, Born, & Wagner, 2010). Given these results as well as the fact that relational inferences from learned premise pairs are significantly boosted after a period of sleep (Ellenbogen, Hu, Payne, Titone, & Walker, 2007), we might observe an even stronger effect in relational lures after sleep.

Conclusion

If entity concepts are the building blocks of cognition, then relations are the mortar that holds them together. Relations provide an organizational structure for semantic knowledge and the ability to abstract information beyond a single learning episode, beyond the entities that instantiated them. Relations are thus the very "fuel and fire of thinking" (Hofstadter & Sander, 2013) and are arguably what makes human cognition so special (Gentner, 2010). The ability to manipulate relations requires that one is able to extract implicit relational information from sensory input, to encode it abstractly and independently from the representations of the entities that instantiate it, and finally, to be able to retrieve it on demand, directly by other instances of the same relation or relational structure. Despite that many models of LTM have largely ignored the mnemonic role of relations. The relational luring effect presented in this study suggests that implicit relations between items are indeed encoded during the initial study of word pairs in an abstract manner such that they subsequently influence associative recognition judgments of different word pairs that share the same relation. This provides evidence that relations have independent abstract representations in long term memory, and that they can be retrieved by different instances. The relational luring effect and its dynamics may be used as benchmarks for extensions of memory models, which cannot explain them with their current specifications. The current results also provide an alternative explanation to the phenomenon of unitization in associative recognition. It also provides additional support for the hypothesis that the constructive nature memory depends not only on semantic, but also on relational similarity. Finally, we suggest that the retrieval of relevant analogues from LTM might occur through accumulation of activation in single relations that are shared between structures in LTM and the structure of the problem at hand.

References

Ahmad, F. N., Fernandes, M., & Hockley, W. E. (2015). Improving associative memory in older adults with unitization. *Aging, Neuropsychology, and Cognition*, *22*(4), 452–472.

Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

Asmuth, J., & Gentner, D. (2016). Relational categories are more mutable than entity categories. *The Quarterly Journal of Experimental Psychology*, 1–19.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*(1), 141.

Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 343.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting Linear Mixed-Effects Models using lme4. *arXiv:1406.5823 [Stat]*.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Blanchette, I., & Dunbar, K. (2002). Representational change and analogy: How analogical inferences alter target representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 672.

Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, *59*(2), 183–199.

Catrambone, R., & Holyoak, K. J. (1989). Overcoming contextual limitations on problem-solving transfer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1147.

Challis, B. H., & Sidhu, R. (1993). Dissociative effect of massed repetition on implicit and explicit measures of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(1), 115–127.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, *92*(2), 149–154.

Cummins, D. D. (1993). “Role of analogical reasoning in the induction of problem categories”: Correction. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 908–908.

Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, *35*(1), 39–49.

Day, S. B., & Goldstone, R. L. (2011). Analogical transfer from a simulated physical system. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(3), 551.

DeWolf, M., Son, J. Y., Bassok, M., & Holyoak, K. J. (2017). Relational Priming Based on a Multiplicative Schema for Whole Numbers and Fractions. *Cognitive Science*. Advanced online publication

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*(1), 1–21.

Diekelmann, S., Born, J., & Wagner, U. (2010). Sleep enhances false memories depending on general memory performance. *Behavioural Brain Research*, *208*(2), 425–429.

Dixon, J. A., & Dohn, M. C. (2003). Redescription disembeds relations: Evidence from relational transfer and use in problem solving. *Memory & Cognition*, *31*(7), 1082–1093.

Doumas, L. A., & Hummel, J. E. (2012). Computational models of higher cognition. *Oxford Handbook of Thinking and Reasoning*, 52–66.

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

Dunbar, K. (2001). The analogical paradox: Why analogy is so easy in naturalistic settings yet so difficult in the psychological laboratory. *The Analogical Mind: Perspectives from Cognitive Science*, 313–334.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. *Proceedings of the National Academy of Sciences*, *104*(18), 7723–7728.

Estes, Z. (2003). Attributive and relational processes in nominal combination. *Journal of Memory and Language*, *48*(2), 304–319.

Estes, Z., & Jones, L. L. (2006). Priming via relational similarity: A copper horse is faster when seen through a glass eye. *Journal of Memory and Language*, *55*(1), 89–101.

Estes, Z., & Jones, L. L. (2009). Integrative priming occurs rapidly and uncontrollably during lexical processing. *Journal of Experimental Psychology: General*, *138*(1), 112–130.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial Intelligence*, *41*(1), 1–63.

Feldman, J. (2013). The neural binding problem(s). *Cognitive Neurodynamics*, *7*(1), 1–11.

Feldman, V., & Kokinov, B. (2009). Analogical Episodes are More Likely to be Blended than Superficially Similar Ones. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum*.

Ferry, A. L., Hespos, S. J., & Gentner, D. (2015). Prelinguistic Relational Concepts: Investigating Analogical Processing in Infants. *Child Development*, *86*(5), 1386–1405.

Forbus, K. D., Gentner, D., & Law, K. (1995). MAC/FAC: A Model of Similarity-Based Retrieval. *Cognitive Science*, *19*(2), 141–205.

Ford, J. H., Verfaellie, M., & Giovanello, K. S. (2010). Neural correlates of familiarity-based associative retrieval. *Neuropsychologia*, *48*(10), 3019–3025.

French, R. M. (2008). Relational priming is to analogy-making as one-ball juggling is to seven-ball juggling. *Behavioral and Brain Sciences*, *31*(4), 386–387.

Gagné, C. L. (2001). Relation and lexical priming during the interpretation of noun–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 236.

Gagné, C. L. (2002). Lexical and Relational Influences on the Processing of Novel Compounds. *Brain and Language*, *81*(1–3), 723–735.

Gagné, C. L., & Shoben, E. J. (1997). Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 71.

Gagné, C. L., Spalding, T. L., & Ji, H. (2005). Re-examining evidence for the use of independent relational representations during conceptual combination. *Journal of Memory and Language*, *53*(3), 445–455.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*(2), 155–170.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752–775.

Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 266–276.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology*, *95*(2), 393.

Gentner, D., & Smith, L. (2012). Analogical reasoning. *Encyclopedia of Human Behavior (2nd Ed.), VS Ramachandran, Ed. Elsevier*, 130–136.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, *12*(3), 306–355.

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

Glass, A. L., Holyoak, K. J., & Kiger, J. I. (1979). Role of antonymy relations in semantic judgments. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 598.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, *33*(2), 1–22.

Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron*, *59*(4), 554–560.

Hilbe, J. M. (2009). *Logistic regression models*. Boca Raton: CRC Press.

Hofstadter, D. (2001). Analogy as the core of cognition. *The Analogical Mind: Perspectives from Cognitive Science*, 499–538.

Hofstadter, D., & Sander, E. (2013). The talent so fundamental that it is fuel to our minds. *New Scientist*, *218*(2915), 30–33.

Holyoak, K. J. (2012). Analogy and relational reasoning. *The Oxford Handbook of Thinking and Reasoning*, 234–259.

Holyoak, K. J., Gentner, D., & Kokinov, B. N. (2001). Introduction: The place of analogy in cognition. *The Analogical Mind: Perspectives from Cognitive Science*, 1–19.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Hristova, P. (2009). Unintentional and unconscious analogies between superficially dissimilar but relationally similar simple structures. In *New Frontiers in analogy research: Proceedings of the second international conference on analogy* (pp. 193–203).

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*(3), 427.

Hummel, J. E., Holyoak, K. J., Green, C., Doumas, L. A., Devnich, D., Kittur, A., & Kalar, D. J. (2004). A solution to the binding problem for compositional connectionism. In *Compositional connectionism in cognitive science: Papers from the AAAI Fall Symposium, ed. SD Levy & R. Gayler* (pp. 31–34).

Jackendoff, R. S. (2002). *Foundations of langage: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Janczura, G. A., & Nelson, D. L. (1999). Concept accessiblity as the determinant of typicality judgments. *The American Journal of Psychology*, *112*(1), 1–19.

Jones, L. L., Estes, Z., & Marsh, R. L. (2008). An asymmetric effect of relational integration on recognition memory. *The Quarterly Journal of Experimental Psychology*, *61*(8), 1169–1176.

Keane, M. (1987). On retrieving analogues when solving problems. *The Quarterly Journal of Experimental Psychology Section A*, *39*(1), 29–41.

Kokinov, B., & French, R. M. (2003). Computational models of analogy-making. *Encyclopedia of Cognitive Science*, *1*, 113–118.

Kokinov, B., & Petrov, A. (2001). Integrating memory and reasoning in analogy-making: The AMBR model. *The Analogical Mind: Perspectives from Cognitive Science*, 59–124.

Kriukova, O., Bridger, E., & Mecklinger, A. (2013). Semantic relations differentially impact associative recognition memory: Electrophysiological evidence. *Brain and Cognition*, *83*(1), 93–103.

Kruschke, J. (2010). *Doing Bayesian data analysis: a tutorial introduction with R*. Academic Press.

Leech, R., Mareschal, D., & Cooper, R. P. (2008). Analogy as relational priming: A developmental and computational perspective on the origins of a complex cognitive skill. *Behavioral and Brain Sciences*, *31*(4).

Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, *11*(3), 126–135.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society*, *7*(3), 283–284.

Montefinese, M., Zannino, G. D., & Ambrosini, E. (2015). Semantic similarity between old and new items produces false alarms in recognition memory. *Psychological Research*, *79*(5), 785–794.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and Their Acquisition*, *1*, 1–55.

Novick, L. R. (1988). Analogical transfer, problem similarity, and expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 510.

Parks, C. M., & Yonelinas, A. P. (2014). The Importance of Unitization for Familiarity-Based Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Pavlova, M., & Kokinov, B. (2014). Analogy Causes Distorted Memory by Blending Memory Episodes. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2729–2734). Austin, TX: Cognitive Science Society.

Payne, J. D., Schacter, D. L., Propper, R., Huang, L.-W., Wamsley, E., Tucker, M. A., … Stickgold, R. (2009). The Role of Sleep in False Memory Formation. *Neurobiology of Learning and Memory*, *92*(3), 327–334.

Perrott, D. A., Genter, D., & Bodenhausen, G. V. (2005). Resistance is futile: The unwitting insertion of analogical inferences in memory. *Psychonomic Bulletin & Review*, *12*(4), 696–702.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427–459.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129.

Popov, V., & Hristova, P. (2014). Automatic Analogical Reasoning Underlies Structural Priming in Comprehension of Ambiguous Sentences. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1192–1197). Austin, TX: Cognitive Science Society.

Popov, V., & Hristova, P. (2015). Unintentional and efficient relational priming. *Memory & Cognition*, *43*(6), 866–878.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93.

R Core Team. (2014). *R: a language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing.

Raffray, C. N., Pickering, M. J., & Branigan, H. P. (2007). Priming the interpretation of noun–noun combinations. *Journal of Memory and Language*, *57*(3), 380–395.

Reagh, Z. M., & Yassa, M. A. (2014). Repetition strengthens target recognition but impairs similar lure discrimination: evidence for trace competition. *Learning & Memory*, 21(7), 342–346.

Reber, T. P., Luechinger, R., Boesiger, P., & Henke, K. (2014). Detecting Analogies Unconsciously. *Frontiers in Behavioral Neuroscience*, *8*.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a Double-Edged Sword: A Computational Model of The Encoding/Retrieval Trade-Off With Familiarity. In *Psychology of Learning and Motivation* (Vol. 48, pp. 271–312). Elsevier.

Reeves, L., & Weisberg, R. W. (1994). The role of content and abstract information in analogical transfer. *Psychological Bulletin*, *115*(3), 381.

Rhodes, S. M., & Donaldson, D. I. (2007). Electrophysiological evidence for the influence of unitization on the processes engaged during episodic retrieval: Enhancing familiarity based remembering. *Neuropsychologia*, *45*(2), 412–424.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*(1), 1–20.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rosch, E. (1999). Principles of categorization. *Concepts: Core Readings*, 189–206.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 491.

Roskies, A. L. (1999). The Binding Problem. *Neuron*, *24*(1), 7–9.

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience*, 3–30.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

Shoben, E. J. (1991). Predicating and nonpredicating combinations. In *The psychology of word meanings* (pp. 117–135). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Silvey, C., Gentner, D., & Goldin-Meadow, S. (2017). *Children's spontaneous comparisons from 26 to 58 months predict performance in verbal and non-verbal analogy tests in 6th grade.* Manuscript submitted for publication.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*(3), 214.

Spellman, B. A., Holyoak, K. J., & Morrison, R. G. (2001). Analogical priming via semantic relations. *Memory & Cognition*, *29*(3), 383–393.

Tulving, E. (1972). Episodic and semantic memory 1. *Organization of Memory. London: Academic*, *381*(e402), 4.

Wisniewski, E. J., & Love, B. C. (1998). Relations versus properties in conceptual combination. *Journal of Memory and Language*, *38*(2), 177–202.

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517.

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2017). The Effects of Probe Similarity on Retrieval and Comparison Processes in Associative Recognition. *Journal of Cognitive Neuroscience*, *29*(2), 352–367

Zheng, Z., Li, J., Xiao, F., Broster, L. S., Jiang, Y., & Xi, M. (2015). The effects of unitization on the contribution of familiarity and recollection processes to associative recognition memory: Evidence from event-related potentials. *International Journal of Psychophysiology*, *95*(3), 355–362.

© 2017, American Psychological Association. Journal of Experimental Psychology: General. DOI: 10.1037/xge0000305

Appendix A: Stimuli used in Experiment 2

Below are the translated exemplars for the 35 relation groups that were used in Experiment 2. The original study and stimuli were in Bulgarian. In Bulgarian none of the exemplars had words that shared their root, while some of the translated version do (e.g. bath and bathroom). In addition, in Bulgarian noun-noun compounds are ungrammatical, and thus these exemplars cannot be perceived as a single linguistic phrase. Some words are repeated in the English translation, but are distinct in Bulgarian (e.g., jar lid, saucepan lid, pot lid). In some exemplars the main relation has been distorted by differences between the languages (e.g., in the exemplars LAYER CAKE, STRATUM EARTH and CRUST PIE, the pie in question is a specific Bulgarian pastry in which layers of crust and cheese are intermixed).

F1, F2, F3 = how many participants generated the exemplar in the 1st, 2nd or 3rd position.

T = total number of participants that generated the exemplar in any position.

SS = sample size

FAN = total number of different responses to the relation.

SPROP = proportion of participants who generated the exemplar

relation	Word 1	Word 2	F1	F2	F3	T	SS	FAN	SPROP
1	book	writer							
1	blueprint	architect							
1	painting	artist	17	7	2	26	59	97	0.441
1	song	composer	4	0	1	5	59	97	0.085
1	poem	poet	2	2	2	6	59	97	0.102
2	wolf	pack							
2	bird	flock							
2	lion	pride	6	2	4	12	60	69	0.2
2	sheep	herd	16	8	2	26	60	69	0.433
2	bee	swarm	2	1	1	4	60	69	0.067
3	eye	vision							
3	ear	hearing							
3	nose	smell	20	16	3	39	64	50	0.609
3	tongue	taste	17	11	7	35	64	50	0.547
3	skin	touch	3	10	5	18	64	50	0.281
4	school	principal							
4	company	manager	7	2	1	10	60	95	0.167
4	orchestra	conductor	1	3	2	6	60	95	0.1
4	university	rector	3	0	2	5	60	95	0.083
4	army	general	2	0	2	4	60	95	0.067
5	umbrella	rain							
5	jacket	cold	4	8	1	13	62	120	0.21
5	sun	glasses	5	3	2	10	62	120	0.161
5	snow	boots	2	0	0	2	62	120	0.032
7	water	pipe							
7	artery	blood							
7	cable	electricity	3	2	2	7	53	94	0.132
7	riverbed	river	4	0	0	4	53	94	0.075
8	bottle	stopper							
8	jar	lid							
8	saucepan	lid	16	7	0	23	58	75	0.397
8	pot	lid	5	3	1	9	58	75	0.155
8	house	roof	5	3	1	9	58	75	0.155
9	floor	building							
9	layer	cake							
9	stratum	earth	2	0	0	2	48	97	0.042
9	crust	pie	2	0	0	2	48	97	0.042
13	cat	paw							
13	donkey	hoof							
13	fish	fin	5	8	5	18	61	65	0.295

relation	Word 1	Word 2	F1	F2	F3	T	SS	FAN	SPROP
13	pigeon	wing	3	8	7	18	61	65	0.295
13	octopus	tentacle	1	4	4	9	61	65	0.148
14	bed	bedroom							
14	bath	bathroom							
14	kitchen	stove	12	7	2	21	61	67	0.344
14	couch	living-room	9	7	3	19	61	67	0.311
14	office	cabinet	3	2	6	11	61	67	0.18
15	sunrise	sunset							
15	birth	death							
15	start	end	22	5	2	29	57	83	0.509
15	day	night	3	4	1	8	57	83	0.14
17	frog	larva							
17	butterfly	caterpillar							
17	seed	plant	0	3	1	4	59	67	0.068
17	embryo	mammal	2	0	1	3	59	67	0.051
19	fruit	pit							
19	planet	core							
19	egg	yolk	9	3	1	13	52	98	0.25
19	walnut	kernel	4	0	0	4	52	98	0.077
25	nurse	hospital							
25	restaurant	waiter							
25	vendor	shop	4	3	5	12	54	92	0.222
25	bank	cashier	2	1	1	4	54	92	0.074
25	mechanic	workshop	2	1	1	4	54	92	0.074
26	chair	furniture							
26	fork	utensils							
26	plate	dishes	4	2	0	6	59	132	0.102
26	shirt	clothes	3	1	0	4	59	132	0.068
26	screwdriver	tools	2	1	1	4	59	132	0.068
27	strike	pain							
27	joke	laugh							
27	insult	crying	3	2	0	5	55	122	0.091
27	kiss	love	2	0	0	2	55	122	0.036
28	telescope	astronomy							
28	microscope	biology							
28	chemistry	chemistry	3	1	3	7	47	103	0.149
28	scalpel	surgery	4	0	1	5	47	103	0.106
28	map	georgraphy	1	0	3	4	47	103	0.085
29	rabies	dog							
29	malaria	mosquito							
29	plague	rat	8	1	0	9	57	107	0.158
29	flu	virus	6	3	1	10	57	107	0.175
31	floor	carpet							
31	table	cloth							
31	window	curtain	9	5	0	14	58	75	0.241
31	mattress	sheet	9	3	1	13	58	75	0.224
31	wall	wallpaper	4	3	2	9	58	75	0.155
33	kilogram	pound							
33	kilometer	mile							
33	centimeter	inch	22	4	0	26	53	56	0.491
33	liter	gallon	5	10	4	19	53	56	0.358
33	Celsius	Fahrenheit	7	4	4	15	53	56	0.283
35	watt	power							
35	length	meter	4	8	0	12	59	71	0.203
35	weigth	gram	3	4	5	12	59	71	0.203
35	voltage	volt	2	2	5	9	59	71	0.153
35	resistance	ohm	1	3	4	8	59	71	0.136
36	garden	fence							
36	membrane	cell							
36	border	state	6	2	1	9	52	96	0.173

 Journal of Experimental Psychology: General. DOI: 10.1037/xge0000305

relation	Word 1	Word 2	F1	F2	F3	T	SS	FAN	SPROP
36	frame	picture	2	2	3	7	52	96	0.135
36	door	frame	9	3	0	12	59	96	0.203
37	minute	clock							
37	month	calendar							
37	second	stopwatch	9	1	1	11	59	92	0.186
37	degree	thermometer	4	5	1	10	59	92	0.169
37	millimeter	line	3	0	1	4	59	92	0.068
40	horse	rider							
40	car	driver	24	6	0	30	59	55	0.508
40	tram	carman	5	6	5	16	59	55	0.271
40	bike	cyclist	4	1	1	6	59	55	0.102
40	motor	rocker	2	1	1	4	59	55	0.068
42	insomnia	caffeine							
42	intoxication	alcohol							
42	satiety	food	4	3	1	8	58	122	0.138
42	euphoria	drug	3	1	0	4	58	122	0.069
43	airplane	airport							
43	ship	port							
43	train	station	13	8	4	25	62	63	0.403
43	car	parking	5	2	0	7	62	63	0.113
43	bus	stop	3	2	2	7	62	63	0.113
46	teacher	student							
46	professor	student							
46	parent	child	6	3	1	10	60	96	0.167
46	master	apprentice	5	1	4	10	60	96	0.167
46	coach	athlete	4	3	2	9	60	96	0.15
47	squirrel	hollow	8	8	2	18	61	77	0.295
47	mole	hole	4	4	1	9	61	77	0.148
47	bear	den	4	4	0	8	61	77	0.131
47	snail	shell	1	3	1	5	61	77	0.082
47	pig	sty	1	2	2	5	61	77	0.082
49	tennis	racquet							
49	baseball	bat							
49	golf	stick	6	3	2	11	60	91	0.183
49	billiard	cue	1	2	2	5	60	91	0.083
49	badminton	racket	2	0	2	4	60	91	0.067
51	sock	leg							
51	glove	hand							
51	hat	head	39	11	2	52	62	72	0.839
51	neck	scarf	0	9	9	18	62	72	0.29
51	bra	breasts	0	3	1	4	62	72	0.065
52	soccer	goal							
52	basketball	basket							
52	volleyball	point	10	1	1	12	49	63	0.245
52	Boxing	knockout	0	6	2	8	49	63	0.163
52	chess	check-mate	3	1	1	5	49	63	0.102
53	necklace	neck							
53	ring	finger	11	7	1	19	63	58	0.302
53	bracelet	wrist	9	6	1	16	63	58	0.254
53	belt	waist	1	5	5	11	63	58	0.175
54	cheetah	savannah							
54	dolphin	sea							
54	monkey	jungle	5	2	1	8	65	107	0.123
54	camel	desert	1	6	2	9	65	107	0.138
54	goal	mountain	0	4	0	4	65	107	0.062
55	bread	dough							
55	cocoa	chocolate							
55	milk	cheese	4	4	3	11	56	101	0.196
55	grape	wine	3	1	4	8	56	101	0.143
55	mince	meatball	1	2	4	7	56	101	0.125

relation	Word 1	Word 2	F1	F2	F3	T	SS	FAN	SPROP
56	tailor	needle	1	1	1	3	66	117	0.045
56	hairdresser	scissors	0	4	1	5	66	117	0.076
56	butcher	knife	2	1	1	4	66	117	0.061
56	fisherman	fishing	2	1	1	4	66	117	0.061
56	lumberjack	ax	2	0	2	4	66	117	0.061

Appendix B: Pretesting of the stimuli for Experiment 2

Method

Participants. A total of 79 participants (58 female) took part in the pretesting. Participants were native Bulgarian speakers that were recruited on social media and were asked in turn to share the study link with their contacts. Their age ranged from 17 to 67 years (M = 38, SD = 12). They were offered a change to win a gift card for finishing the full study and for giving very typical exemplars of a relation. They were instructed that the two people, who gave responses that where shared with the most other people would win gift cards valued at 17$ and 12$.

Materials and procedure. We selected 2 exemplars for each of 58 relations, which we had already determined to be dominant exemplars of their relation in a previous pretesting study (for details, see Popov & Hristova, 2015). Materials were administered through an online survey platform (http://esurv.org). For each relation the two exemplars were presented together and we asked participants to generate up to 3 novel exemplars for each relation. For example, participants where given the following two word pairs:

NURSE HOSPITAL

WAITER RESTAURANT

They were told that the two word pairs are analogical, because they share the same relation. Namely, a "nurse' works in a "hospital", just as a "waiter" works in a "restaurant". They were told that this is not a creativity test and that they should attempt to write down the first analogical word pair that comes to mind. The whole procedure took between 30 and 120 minutes, and participants were told that if they are having trouble with some examples, it is better to move on to the next and that even partial responses will be of use. Relations were presented in random order for each participant, thus even if participants gave up before completing the task, responses were equally spread among all relations.

Results and stimuli selection.

All responses were subsequently spell-checked and manually inspected before analysis, to remove differences between singular and plural forms, alternative spellings, etc. For each relation we counted the proportion of participants who gave each exemplar in each position (first, second or third response). Each of the 58 relations was given at least one answer from 47 to 67 participants. Each relation received between 117 and 186 separate responses. For each relation group, we first identified the 8 most dominant responses. Going for the least to the most dominant response across relations, we manually and iteratively removed exemplars that shared one or two of their words with exemplars to other relations, until only responses with unique words remained. Only 35 of the initial 58 relations had more than 1 generated exemplar left, and for each we selected the 2 (10 relations) or 3 (25 relations) most dominant responses. Combined with the original two exemplars, this resulted in 25 relations with 5 exemplars and 10 relations with 4 exemplars.

---

i We were not able to generate 5 exemplars for all relations.

ii Though often used interchangeably, they are conceptually and mathematically distinct (Hilbe, 2009)

iii Since "recombined" was the reference level in the multinomial model, the RR for comparing non-reference levels can be obtained by dividing the RR for "old" vs "recombined" by the RR for "new" vs "recombined". Specifically $\frac{P(\text{old})}{P(recombined)} = \exp\left(\frac{log\left(\frac{P(old)}{p(recombined)}\right)}{log\left(\frac{p(new)}{p(recombined)}\right)}\right)$

iv All reported changes are multiplicative. For example, by the final trial participants were $3.2 * 1.005^{527} = 3.2 * 0.34 = 44.3$ times more likely to give an "old" rather than a "recombined" response to intact pairs.