Evidence of analogical re-representation from a change detection task

Daniel C. Silliman*, Kenneth J. Kurtz

Binghamton University, United States

1. Introduction

The richness and power of human cognition derives in part from the ability to experience analogical matches across space, time, and settings that are grounded in common relational structure despite differing superficial characteristics (Gentner, 1983, 2003). Consider, for example, seeing an animal that is highly fit to its environment and being reminded of a thriving business. The business and animal share little physical resemblance, yet they are similar to the extent that they both embody principles of natural selection. The customers’ tastes and environmental hazards both serve as selective pressures for which the business and animal are both well suited. Our capacity to notice this kind of deep relational similarity and use it to inform our actions is central to research on analogical reasoning. Though work inspired by theories and process models has revealed much about the nature of relational similarity (Gentner, Holyoak, & Kokinov, 2001; Vendetti, Matlen, Richland, & Bunge, 2015), certain important theoretical assumptions remain empirically impoverished. The present work concerns one such assumption on the mechanisms that enable the rich flexibility inherent in human analogy-making.

The problem of flexibility is best understood within the context of the structure mapping theory (SMT) of analogy (Gentner, 1983, 1989). According to SMT, analogy is defined as a mapping of relational structure between two representations in working memory, where representations are assumed to take the form of propositions. Propositions as structured representations consist of a symbolic vocabulary of relations, objects, and attributes. For convenience, parallels may be drawn between representational elements and parts of speech. Relations may be understood as tantamount to adjectives, while objects are tantamount to entity nouns, and attributes are tantamount to adjectives.

An analogy is attained when the mapping of two structured representations reveals conceptually similar relations (Gentner & Kurtz, 2006)—that is, the relations between objects are the same across representations. The following examples serve to illustrate this definition:

(1) During the evening, many people tend to browse the internet all at once. The resulting bandwidth congestion prolongs packet latency, making everybody wait even longer.

(2) When the work day ends, everyone attempts to drive on the freeway at the same time. The ensuing gridlock slows all road traffic so that nobody can get home on time.

The above scenarios both depict how many people accessing a resource at the same time can impede access to the resources. The objects (e.g., people, internet, freeway) are similar insofar as they play the same role in the relational structure. This common relational structure causally links the initial relations (browse, drive) to the latter relations (prolongs, slows), and it is this similarity between higher-order relations—that is quintessential to perceiving analogy (Gentner, 1983). SMT refers to the preference for matching higher-order relations as systematicity (Clement & Gentner, 2006).
In the above examples, it is not strictly the common abstract causal relation that makes the two situations appear analogous. There are many examples of causal relationships that are not analogous with the examples provided. The conceptual content of the arguments to the higher-order relation matter as much as the relation itself (Gentner & Kurtz, 2006; Gentner & Markman, 2006).

The above examples also capture the issue critical to the present work. Most readers are able to recognize the relational similarity between the two scenarios despite the lower-order relations not being conceptually synonymous. The relations access and drive are not synonymous (neither are prolong and slow), yet still contribute to the relational similarity that is crucial to analogy. This observation may appear trivial *prima facie*, but it represents a core issue that any theory must address for a complete account of analogy. The crux of the issue lies with defining a similarity criterion that is strict enough to exclude arbitrary matches, yet not so strict as to exclude potentially meaningful matches. Arbitrary matches become especially problematic if one considers the set size of all possible matches in long-term memory. SMT reconciles the tension between these opposing pressures by suggesting that mappings abide by the constraint of *tiered identity* (Forbus, Ferguson, Lovett, & Gentner, 2017; Gentner & Forbus, 2011; Markman & Gentner, 2000). This constraint suggests that when engaging in analogy, only conceptually identical (i.e., synonymous) relations are mapped by default. Tiered identity allows for initially non-identical matches if, (a) the relations are the arguments of matching higher-order relations (Markman & Gentner, 2000); or (b) one or both relations can be transformed to attain partial identity (Gentner & Forbus, 2011; Gentner & Kurtz, 2006). Satisfying the latter condition requires re-representation of relational predicates so an underlying conceptual commonality can be matched—i.e., partial identity (Gentner & Wolff, 2000; Kotovsky & Gentner, 1996; Yan, Forbus, & Gentner, 2003). Re-representation is typically characterized as a process of representing similar words in terms of structured sub-components (Gentner & Kurtz, 2006; Kurtz, 2005). These semantic sub-components are then compared to reveal potential identity matches. For example, the relations of drive on(road) and browse(internet) can be re-represented to include content overlap in terms of making use of a shared resource and traversing a space (physical or cyber). This process does not require a substitution of lexical representation, but rather an elaboration of internal description. Given that the new description is semantically broadened to subsume the meaning of its constituent elements, the re-representation can be construed as an abstraction of the original content. It follows that if no commonalities are present, the re-representation does not take place and the relational predicates fail to be mapped.

Re-representation is the key mechanism allowing tiered identity to address the problem of arbitrary matches while still maintaining the flexibility evident in human analogical reasoning. Despite the critical importance of re-representation to SMT, very little work has demonstrated re-representation in adult behavioral data (Day & Asmuth, 2017; Gentner & Kurtz, 2006; Kurtz, 2005). The lack of studies is certainly tied to the inherent difficulty of evidencing a construct like re-representation. Indexing the psychological representation of a stimulus presents considerable challenges; tracking changes in representation over time only compounds these difficulties. The existing studies address these challenges in different ways.

Kurtz (2005) evidenced re-representation by prompting participants to compare images with a shared relation (e.g., bottles and pens both contain something). Notably, this relation was always less salient in one of the items. After comparison, participants were probed as to whether the relation applied to the less-salient item. Relative to non-comparison groups (e.g., rating the typicality of a pen), the relational comparison group affirmed the relation more often. Gentner and Kurtz (2006) had participants determine the analogical similarity of sentence pairs that varied in the conceptual similarity of their relations (i.e., synonymous, similar, dissimilar). Their reaction time (RT) data revealed that synonymous relations resulted in fast determination of analogy, dissimilar relations resulted in fast determination that a pairing was not analogous, but similar relations resulted in slow and mixed responses. This behavior was interpreted as evidence that participants were attempting to re-represent similar but non-identical relations, and this extra processing consequently prolonged response latency. Day and Asmuth (2017) had participants evaluate similarity between a base and target sentence with a common relational structure. Participants then evaluated a subsequent pairing between the same base and a different target sentence. Critically, depending on condition, this subsequent comparison either featured the same relational commonality as the first comparison, or a different but equally valid relational commonality. When the second comparison featured the same relational structure as in the first, the sentences were rated as more similar to each other. Day and Asmuth reasoned that participants’ initial comparison re-represented the construal of the base sentence such that it was perceived as less similar to a subsequent sentence with a different relational structure.

Among this evidence, there is nothing that directly shows representational change at the level of the individual relational predicate where re-representation is predicted to occur. Instead, the extant studies infer re-representation from changes in participants’ evaluation of whole sentences or objects (relative to a control) based on comparison to a similar item intended to prompt re-representation. Further, there is no study to our knowledge that has been able to demonstrate a link between judgements of analogical similarity and the likelihood of re-representation—a connection that must exist given the theoretical treatment of re-representation.

The present status of re-representation is that it is: (1) vital in theoretical terms to a leading and highly successful psychological account of analogy (SMT); and (2) only minimally evidenced as an actual psychological mechanism. In the absence of a broad empirical foundation, some researchers have voiced doubts about the psychological reality of re-representation (Ramscar & Yarlett, 2003) and others have raised concerns about the computational tractability of the mechanism as currently proposed (Wareham, Robere, & van Rooij, 2012). In addition, there exists computational models that can account for analogical phenomena without relying on tiered identity (Holyoak & Thagard, 1989; Ramscar & Yarlett, 2003). LISA (Hummel & Holyoak, 1997), a leading competitor to the structure-mapping theory of analogy, avoids the issue of tiered identity by representing relations (and objects) via compositional activation units in a symbolic-connectionist architecture. The lowest layer of these units is responsible for capturing semantic primitives. For LISA, an analogy involving similar but non-identical relations would entail overlapping activations of many of the same semantic primitives across both base and target representations. Consequently, partial identity is explicitly determined during every analogical evaluation. Assuming distributed semantic representation by default means that LISA has no use for re-representation. Although the process of reaching an alignment differs, both accounts rely on being able to put similar but non-identical relations into correspondence. SMT differs in assuming a specialized process of re-representation that can be triggered in order to permit such matches (while otherwise maintaining a strict requirement of identity of representational predicates).

Given the robust evidence that has accrued for nearly all other aspects of SMT, it comes as a surprise that the theory still must carry a promissory note with regard to re-representation. Without sufficient empirical grounding of re-representation, structure mapping theory could be seen as having the proverbial Achilles’ heel. We know of no other proposals that allow SMT the required flexibility in mapping granted by re-representation. The primary aim of the present work is to address this empirical lacuna by using a novel memory-based, change detection paradigm designed to, (a) test for representational change at the level of individual relational predicates; and (b) test for a connection between user-generated ratings of analogical similarity and the
The critical prediction is this: if participants performing an analogical evaluation attempt to re-represent conceptually similar, but not synonymous relations, the resulting re-representation should leave them less likely to detect a lexical change that parallels the re-representation (same underlying meaning) in the recognition test. These lexical lures specifically capture the partial identity between base and target relations. Fewer changes to relations should be detected from analogical pairings than non-analogical pairings since the re-representation process is invoked to successfully process analogical matches. This prediction follows from prior work by Johnson, Bransford, and Solomon (1973) that suggests that individuals often falsely recognize new information that is broadly consistent with information they had previously been exposed to. Consider the two examples from earlier:

(1) During the evening, many people tend to browse the internet all at once. The resulting bandwidth congestion prolongs packet latency, making everybody wait even longer.

(2) When the work day ends, everyone attempts to drive on the freeway at the same time. The ensuing gridlock slows all road traffic so that nobody can get home on time.

If participants were to attempt re-representation to resolve this analogy, the representations of browse and drive on would change to reflect the partial identity of make use of. Consequently—compared to participants who did not see the base passage in the context of an analogy—the participants who viewed the earlier analogy would be less likely to detect a change from browse to use. It follows that: (1) higher analogical similarity should predict fewer detected changes in relational content and (2) only relations should show the effect since detection of changes made to objects or attributes should not be significantly affected by analogical evaluation.

2.1. Methods

2.1.1. Participants

Psychology undergraduates at Binghamton University (n = 58) participated in this study in exchange for partial credit toward a course requirement. The sample size was based on similar studies in the related literature (e.g., Day & Asmuth, 2017). All participants gave informed consent and were debriefed in accord with the policies of the SUNY Binghamton IRB.

2.1.2. Materials and design

The materials for the analogical evaluation phase consisted of text passages in analogical or non-analogical pairs. Each base passage had a corresponding analogous and non-analogous target passage that it could be paired with (see Table 1 for an example; see the Appendix for a report of all key relations, relational lures, as well as concreteness (Brysbaert, Warriner, & Kuperman, 2014) and semantic relatedness (Bird, Klein, & Loper, 2009; WordNet, 2010) ratings). Pilot testing was conducted to develop passages that would be seen by participants as analogical or non-analogical. Two different forms were used to counterbalance which target passage appeared with a given base, with no base passage being repeated in the rating phase. Each form contained a

### Table 1

<table>
<thead>
<tr>
<th>Base</th>
<th>Analog</th>
<th>Non-Analog</th>
</tr>
</thead>
<tbody>
<tr>
<td>The veteran shopper was at a swap-meet stall, hoping to purchase rare computer parts. Only astute shoppers with a keen eye could get the most sought after equipment. Even though many other buyers were looking for the rare parts, the veteran shopper got ahold of them first. Only junk was left by the time he was done.</td>
<td>The grey vulture circled the desert, interested in foraging for scraps. Only the vultures with a sharp sense of smell could find the most recently passed animals. Even with all the other vultures scavenging for the same thing, the grey vulture found the carcass first. There were only bones remaining after it finished.</td>
<td>The art critic went down to the new museum to appraise a gallery that had recently opened up. Because she carelessly left her VIP badge at home, she had to endure standing in line with the general public. The lengthy delay helped the art critic to really appreciate the gallery once she finally managed to get inside.</td>
</tr>
</tbody>
</table>
Table 2
Sample passage featured in change detection. Changes bolded.

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Paired with analog</th>
<th>Paired with non-Analog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change</td>
<td>The veteran shopper was at a swap-meet stall, hoping to</td>
<td>The veteran shopper was at a swap-meet stall, hoping to</td>
<td>The veteran shopper was at a swap-meet stall, hoping to</td>
</tr>
<tr>
<td></td>
<td>purchase discarded electronics. Only astute shoppers</td>
<td>salvage discarded electronics. Only opportunistic</td>
<td>examine discarded electronics. Only patient</td>
</tr>
<tr>
<td></td>
<td>with a keen eye could get the most sought after equipment. Even though</td>
<td>shoppers with a keen eye could get the most sought after</td>
<td>shoppers with a keen eye could get the most sought after</td>
</tr>
<tr>
<td></td>
<td>many other buyers were looking for the spare parts, the veteran shopper</td>
<td>hardware. Even though many other buyers were hunting</td>
<td>hardware. Even though many other buyers were</td>
</tr>
<tr>
<td></td>
<td>got hold of them first. Only junk was left by the time he</td>
<td>for the same spare pieces, the veteran shopper got hold of</td>
<td>were waiting for the same spare pieces, the veteran shopper</td>
</tr>
<tr>
<td></td>
<td>was done</td>
<td>them first. Only junk was left by the time he was done</td>
<td>got hold of them first. Only junk was left by the time he was done</td>
</tr>
</tbody>
</table>

total of five passage pairs; the first and last were always fixed filler passages that were not evaluated during scoring. The purpose of the filler passages was to mitigate any serial-position effects (Neath, 1993) on critical passages due to the memory-based measure. The ordering of the three critical passage pairs was randomized across forms.

The change detection task featured the critical base passages from the evaluation phase with strategic modifications: two relations, two objects, and one attribute were changed in each passage (see Table 2 for sample materials). Each sentence could feature multiple changes to components or no changes at all. To reduce the possibility of participants noticing this pattern, one of the original filler passages was included with no changes. The relations were replaced with a word that the experimenters believed best captured the underlying meaning between the corresponding base and target relations as they appeared in the evaluation phase. To help ensure that potential differences in lure detection rates were not attributable solely to gist encoding (Bransford & Franks, 1971; Wolfe, Reyna, & Brainerd, 2005), changed relations from non-analogical pairings also reflected the word that best captured the underlying meaning of the target relation in the earlier phase. Any gist processing or relational priming should occur equally for base passages following analogical or non-analogical comparison. The changes made for the objects and attributes also maintained the general conceptual content of the words originally used in the base and target. As mentioned earlier, the SMT prediction concerns re-representation of relational content. Evidence from the detection rates for objects or attributes could indicate some type of effect of analogical comparison but would not serve as good evidence for re-representation.

2.1.3. Procedure
Research assistants ran participants individually in isolated rooms. The research assistants first gave participants the analogical evaluation phase on a paper handout. Participants were allowed as much time as needed to complete this phase and were allowed to flip back to earlier pages in the packet. Instructions preceding every pairing asked participants to rate how analogous the pairings appeared using the provided Likert scale (1 – 7). The instructions did not provide a definition of analogy to the participants (as doing so might bias them to attend particularly to the relations and potentially confound the results). When the participants completed the analogical evaluation phase, the research assistants replaced the paper handout with a handout for the detection phase. At no point in the evaluation phase did the assistants or instructions indicate to the participants that they would be engaging in a later change detection phase. The one filler passage included in the change detection phase was unchanged from its original form and always took the position of the second passage; the critical passages were randomly ordered in the remaining positions. The instructions at the top of the change detection sheet told participants that they would be seeing some of the passages from earlier and their task was to indicate any changes in the passages by circling or underlining the word(s) that might have changed. The instructions suggested that passages may or may not feature changes. A successfully detected change was coded as a hit, while undetected changes were coded as a miss. Indication that an unchanged word had changed was coded as a false alarm. As with the evaluation phase, participants were allotted as much time as required to complete this portion of the experiment. The majority of participants concluded the entirety of the experiment within 15 min.

2.2. Results and discussion
Participants rated the analogous passages as analogous ($M = 6.16$, $SE = 0.1$) and the non-analogous passages as non-analogous ($M = 2.51$, $SE = 0.17$). Our primary prediction was that base passages seen as analogous in the evaluation phase would result in fewer changed relations detected than passages seen as non-analogous. A cumulative link mixed model (Christensen, 2015) was built using the R programming environment (R Core Team, 2016) to test this relationship. Number of changed relations detected was predicted with participant and base passage treated as random effects, while the number of prior analogies seen in the evaluation phase, number of false alarms per trial, and pairing type were treated as fixed effects. Number of prior analogies was included as a fixed effect because the counterbalancing resulted in one form containing more analogies than the other (including the fillers). The cumulative link mixed model (CLMM) affirmed our predictions (see Fig. 1), revealing that bases paired with analogous resulted in significantly fewer changed relations detected than bases paired with non-analogs ($β = -0.849$, $SE = 0.323$, $p = 0.009$). Number of priors seen ($β = -0.122$, $SE = 0.304$, $p = 0.688$) and false alarms ($β = -0.017$, $SE = 0.04$, $p = 0.674$) did not reveal significant differences. A second CLMM was built to test whether higher ratings would be associated with fewer changed relations detected. The second model was the same as the first with the exception of the aforementioned switch in fixed effect. The model affirmed that higher analogical ratings were associated with fewer changed relations detected ($β = -0.16$, $SE = 0.073$, $p = 0.028$). Significance of the number of priors seen ($β = -0.123$, $SE = 0.073$, $p = 0.028$).
SE = 0.298, p = 0.682) and false alarms (β = -0.014, SE = 0.04, p = 0.725) did not change.

Our ancillary hypothesis predicted that there would be no differences in change detection rates for either objects (nouns) or attributes (adjectives) between passages that were either analogous or non-analogous. The data for object change detection indicate a reversal of the pattern seen for relations, while the data for attribute change detection suggest more modest differences between pairing types (see Fig. 1). Because we predicted null differences, a Bayes Factor (BF) was calculated (Morey & Rouder, 2018) for both object and attribute change detection (compared across passage pairing types) to better interpret evidence for or against the null (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The BF revealed the evidence for the null hypothesis was 1.2 times more likely for objects, and 5.14 for attributes. Common interpretations of BFs (Kass & Raftery, 1995) suggest these values are respectively ‘weak’ and ‘substantial’ evidence for the null. In light of the unexpectedly weak evidence for the null with regard to object change detection an exploratory CLMM model was built to assess the alternate hypothesis. Object change detection was predicted with the same fixed and random effects as in the condition-based model predicting relations detected. This model revealed that changed objects were more likely to be detected following analogical evaluation (β = 0.843, SE = 0.345, p = 0.015). Though number of priors seen did not significantly predict changed object detection (β = 0.179, SE = 0.342, p = 0.6), a higher number of false alarms was significantly associated with a greater probability of changed object detection (β = 0.129, SE = 0.051, p = 0.012). Recall that the relation changes were found to be less detectable for base passages in analogical comparisons; here we see that object changes were more detectable when the passage was seen as participating in an analogy.

This finding, while initially unexpected, can be explained by related literature on alignable differences (Gentner & Markman, 1994; Markman & Gentner, 1993). Research conducted in this area has revealed that similarity comparisons highlight differences between representations that are structurally linked to their commonalities. These alignable differences are naturally more salient than difference not connected to any structural commonalities (Gentner & Markman, 1997). In the present materials, some objects of analogous passages can be considered alignable differences since they are the arguments of the common lower-order relational predicates. By contrast, objects in non-analogous passages, while still the arguments of relational predicates, are not connected by a common relational structure, and are therefore considered non-alignable differences. The more salient alignable differences in the analogical passages should be expected to draw more attention, consequently affecting encoding and later retrieval during the test phase. These findings may also be related to the levels of processing literature (Craik & Tulving, 1975)—specifically, the compatibility effect. Rating a comparison as particularly analogous can be interpreted as a positive response. Research from the levels of processing literature would suggest that answering in a more confirmatory manner should enhance memory for those items (in our case, objects). While this is an issue of interest, the question of whether the object detection findings are best understood in terms of alignable difference or levels of processing is not central to the present investigative goals.

The significant relationship between number of false alarms and changed objects detected was also not predicted. This finding potentially reflects participants with a higher response bias. Such participants would require comparatively less evidence to indicate that a change has been made. Unfortunately, a measure of bias and sensitivity cannot be conducted on these data since each trial presents multiple and non-uniformly distributed opportunities for false alarms, hits, misses, and correct rejections. As for why this same relationship is not observed for relational information, it may be that the relations are just generally less stable in memory (Asmuth & Gentner, 2017). Notably, this difference in mnemonic stability would apply to all relations, regardless of whether they were featured in analogous or non-analogous pairings.

These results demonstrate strong empirical evidence for re-representation as predicted by SMT. When tasked with comparing pairs of passages, participants who evaluated an analogous pair were more susceptible to a later lexical lure that captured the underlying meaning of the passages’ corresponding relations. These same analogical evaluations did not influence a further susceptibility to lures that took the form of attributes and had the opposite effect for objects. For relations and attributes, the pattern of participant behavior was consistent with the claims of SMT. The findings concerning objects, while not initially predicted, are not problematic to our primary prediction regarding re-representation. This evidence contributes notably to the otherwise sparse empirical foundation of re-representation. Unlike prior approaches, the change detection paradigm reflects the actual change in representation where it is predicted to occur. The lexical-level measure exists on the scale of the targeted propositional element, instead of whole sentences or representations. While propositional elements are not theoretically tantamount to their lexical counterparts, evidence at this scale is arguably the most precise behavioral evidence of this phenomenon. Another strength of the present findings is that the analogical similarity ratings generated by the participants were used to predict the re-representations. Though prior studies provide data on similarity ratings, none had statistically leveraged these ratings to evidence their measures of re-representation.

One limitation of the present findings warrants discussion. The changed relations in the base passages at test were different depending on the pairing type. This leaves open the possibility that the replacements in the base following analogical pairing were simply a better match than the replacements in the base following non-analogical pairing; however, it is unlikely that the present findings were driven entirely by this coincidence across all twelve possible changed relations for the majority of participants.

Nonetheless, steps are taken in Experiment 2 to address this concern. At present, the best explanation for the given pattern of behavior is that participants evaluating two passages with non-identical relational predicates as analogous invoke re-representation in order to bring the relational content of the passages into identity-based alignment, and as a result become less likely to detect changes to the passage that made it more like the re-represented form than the original form. Having established the change detection paradigm as an effective means of demonstrating re-representation, the goal of Experiment 2 was to critically examine how conceptual similarity of lower-order relations impacts re-representation and overall analogical similarity.

3. Experiment 2

The evidence from Experiment 1 suggests that re-representation arises in the course of analogical comparison. A comprehensive investigation of re-representation, however, involves not only demonstrating that it occurs when predicted, but also demonstrating that it does not occur when not predicted. Experiment 1 had revealed less re-representation in passage pairs designed to appear non-analogous. The aim of Experiment 2 was to investigate how conceptual similarity (or lack thereof) between lower-order relations impacts re-representation. Prior work by Gentner and Kurtz (2006) suggests that when lower-order relations in sentence pairs are too conceptually dissimilar, participants are much less likely to judge the pair as analogous. This pattern of findings is consistent with the SMT claim that relations that are too conceptually dissimilar cannot be re-represented to attain partial identity. Failed re-representation may even result in the isolated relations being judged as even more dissimilar than before the comparison (Boroditsky, 2007).

The goals of Experiment 2 were to: (1) directly replicate the main finding of Experiment 1 using a new set of materials; and (2) to investigate the role of lower-order relational similarity in re-representation. Toward the latter goal, an additional passage type was developed that featured common higher-order relations with conceptually
dissimilar lower-order relations (referred to as a false analogy, see Table 3 for an example). The matching higher-order relational structure should prompt a failed attempt at re-representation in the conceptually dissimilar lower-order relational predicates. The failure to re-represent should negatively impact the analogical similarity score such that the ratings for false analogies are significantly lower than ratings for true analogies. False analogies will most likely not be rated as low as non-analogous comparisons, as they still share a common higher-order relation between passages. Since false analogies do not result in re-representation, we expected significantly better change detection performance (higher rate of detection) for false analogies than for true analogies.

3.1. Methods

3.1.1. Participants

Psychology undergraduates at Binghamton University (n = 90) participated in this study in exchange for partial credit toward a course requirement. All participants gave informed consent and were debriefed in accord with the policies of the SUNY Binghamton IRB.

3.1.2. Materials and design

New passages were designed for all comparison types to extend evidence of re-representation to a different set of materials. These materials featured three bases with three possible analogous targets (true analogies), three non-analogous targets, and three false analogy targets. In addition to providing convergent evidence and generalizability, the new materials were designed to correct potentially problematic issues with the prior set. The key lower-order relations for both true analogies and non-analogies in the rating phase were now the same (see Table 3). Due to the nature of the question under investigation, the key relation for false analogies was necessarily different from both true and non-analogies.

The lexical lures in the first experiment were generated via experimenter intuition. While this is not necessarily problematic, it may not be the best way to capture the change in representation for most participants. To improve upon this, the relational lures in the present experiment were generated by an independent group of participants in a norming study. The norming participants were given the passage pairs with the critical relations in bold and were asked to write down as many words as they could think of that could replace both the bold words and still preserve the meaning of both sentences. This task was conducted between-subjects with all the comparison types. Possible replacement words were tallied, candidate replacements that were deemed too synonymous with one of the two words were removed, and the final lexical lure for each target pairing was chosen according to whichever of the remaining words had the highest tally. Non-relational elements were left unchanged since the goals of the present experiment only concerned relations.

Despite using this method for generating the lures, the base passages for true and non-analogies featured the same lure during change detection (see Table 4). This ensured that any differences in change detection between analogous and non-analogous pairings was not due to certain changes better capturing the meaning of their replacements than others. Since the false analogy comparisons necessarily had to feature conceptually dissimilar lower-order relations, the resulting lexical lure generated by the norming participants was often different from the lure used for the other two comparison types. Designing target passages for the false analogies required the lower-order relations to be sufficiently conceptually dissimilar while still allowing for a viable lexical lure for the change detection phase.

3.1.3. Procedure

Relatively few changes were made to the procedure used in Experiment 1 in order to achieve the most direct replication. Due to the inclusion of the new false analogy condition, however, three counterbalanced forms were used instead of two. For the three base passages that would re-appear in the detection phase, one was paired with a true analogy, one was paired with a non-analogy, and one was paired with a false analogy. As with Experiment 1, a participant never saw a base passage paired with more than one target passage and all key comparisons were bookended by two filler passages.

3.2. Results and discussion

True analogies were predicted to result in the highest analogy rating with false analogy and non-analogies resulting in significantly lower ratings. The true analogy (M = 5.14, SE = 0.17) and non-analogy (M = 3.03, SE = 0.16) comparisons partially support this prediction and follow from the findings in Experiment 1; the false analogy comparison (M = 4.95, SE = 0.18), however, appears to be very close in ratings to the true analogues. A CLMM (Christensen, 2015) was built predicting rating with comparison type as a fixed effect, and participant and base passage as random effects. The model revealed that non-analogues resulted in significantly lower ratings than true analogies (β = -2.538, SE = 0.321, p < 0.001), while false analogies did not statistically differ from true analogues (β = -0.303, SE = 0.274, p = 0.27). Non-analogues were also found to result in significantly lower ratings than false analogues (β = -2.235, SE = 0.303, p < 0.001). The model affirms that participants viewed the true and false analogy comparisons as more analogical than the non-analogical comparison.

False analogies were predicted to result in significantly more changes detected than true analogies. The rationale was that conceptually dissimilar lower-order relations should be difficult to re-represent, thus leaving participants less susceptible to the lexical lure. The data partially support this prediction. The mean proportion of relations detected for true analogies (M = 0.43, 95% CI [0.35, 0.54]) is smaller than the proportion detected for non-analogues (M = 0.62, 95% CI [0.52, 0.72]); false analogues, however, appear to result in detection rates somewhere in-between the other two comparison types (M = 0.51, 95% CI [0.41, 0.61]). A logistic mixed-effect model (Bates, Maechler, & Bolker, 2015) was built with the same predictors as the CLMM, as well as the number of false alarms as a fixed effect. Relative to true analogues, detection rates following non-analogous comparisons were significantly higher (β = 1, SE = 0.349, p = 0.004), while detection rates following false analogues were not significantly different (β = 0.628, SE = 0.345, p = 0.068). Relative to false analogues, detection rates following non-analogous comparisons were not significantly different (β = 0.379, SE = 0.340, p = 0.266). False alarm rates were not found to significantly predict detection rates (β = -0.06, SE = 0.06, p = 0.318).

The lack of significant differences between false and true analogies may indicate that some non-negligible portion of participants had successfully re-represented the relations in such a way as to leave them vulnerable to the lexical lure. Notably, change detection rates for false analogies also were not significantly different than the rate associated with non-analogies, suggesting that these relations may have been difficult to re-represent for some, but not for others.

In order to more fully relate these change detection results to Experiment 1, a rating-based analysis was also conducted. The former logistic mixed-effect model was used with rating in place of condition. The model revealed that rating no longer significantly predicts change detection (β = -0.03, SE = 0.781, p = 0.695). False alarms did not significantly predict change detection under this model either (β = -0.06, SE = 0.06, p = 0.318).

Note: The output of fixed effects changes slightly when re-leveling other factors for comparison (e.g., comparing false and true analogies to non-analogies vs. non-analogies and true analogies to false analogues). The other results for false alarm rates are not reported as the findings regarding significance does not change.
0.045, $SE = 0.059$, $p = 0.446$). Consequently, the evidence for replication of Experiment 1 findings with Experiment 2 materials is somewhat mixed. While the condition-based analysis replicated the condition-based analysis of Experiment 1, the ratings-based analysis did not.

The difference between these analyses is whether they are based on what the experimenters consider to be analogous versus what participants reported as analogous. While it is tempting to value the latter more highly as it speaks to the actual experience of each individual, it is important to consider the vagaries of self-report including the quality of participants’ understanding of analogy and degree of insight into the extent that they are experiencing it. In Experiment 1, there was a high degree of agreement between what the experimenters considered analogous and what the participants reported. Unfortunately, this convergence was considerably diminished with the different materials (and participants) of Experiment 2. This was reflected in a higher rate of intermediate-level scores (corresponding to an evaluation of the pair as somewhat analogous) for Experiment 2 (see Fig. 2). In terms of the overall scores, mean ratings for analogical pairings dropped from Experiment 1 ($M = 6.16, SE = 0.1$) to Experiment 2 ($M = 5.14, SE = 0.17$) and the mean ratings for non-analogous pairings went up from Experiment 1 ($M = 2.51, SE = 0.17$) to Experiment 2 ($M = 3.03, SE = 0.16$).

It is relevant to note that the reported Likert ratings may simply reflect the participants’ confidence in a decision, rather than how analogous a comparison is. Under this interpretation, whether two things are considered analogous (a binary evaluation) dissociates from the reported rating. Between the change in materials and lack of instruction as to what constitutes an analogy, participants may ultimately be construing situations as analogous or non-analogous but providing less confident reports of their decision. As such, we could expect a replication of the conditions-based model but not necessarily the ratings-based model.

In light of the partial inconsistency between Experiments 1 and 2, two exploratory CLMMs were built to examine rating and condition as predictors of change detection using data pooled from both experiments. The data included the relations detected for both conditions in Experiment 1 and the same conditions in Experiment 2 (excluding false

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Sample materials from Experiment 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base</td>
<td>True Analogy</td>
</tr>
<tr>
<td>The farmer <em>smothered</em> his crops with new, untested pesticides. After the change in pesticides, nobody in the market wanted to buy his vegetables</td>
<td>The apprentice carpenter <em>layered</em> his furniture with a cheaper lacquer than usual. His mentor wouldn’t accept the dip in quality and rejected the piece</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Sample passage featured in change detection. Key relational change bolded.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>True Analogy/Non-Analogy</td>
</tr>
<tr>
<td>The farmer <em>smothered</em> his crops with new, untested pesticides. After the change in pesticides, nobody in the market wanted to buy his vegetables</td>
<td>The farmer <em>covered</em> his crops with new, untested pesticides. After the change in pesticides, nobody in the market wanted to buy his vegetables</td>
</tr>
</tbody>
</table>

Fig. 2. Distribution of rating scores by condition for Experiments 1 and 2.
analogy). The CLMMs included participant and base passage as random effects, with number of false alarms, experiment, and either rating or condition as fixed effect. The model with rating as predictor revealed that higher ratings did significantly predict lower rates of change detection ($\beta = -0.209$, $SE = 0.078$, $p = 0.008$). Additionally, the model revealed a main effect of experiment such that Experiment 2 evidenced lower rates of change detection overall ($\beta = -1.59$, $SE = 0.711$, $p = 0.025$). Number of false alarms was not significantly predictive of changed relations detected ($\beta = -0.026$, $SE = 0.034$, $p = 0.45$). Finally, no interaction was evidenced between experiment and rating ($\beta = 0.185$, $SE = 0.106$, $p = 0.081$).

The model with condition as predictor revealed that analogous pairings predicted lower rates of change detection than non-analogous pairings ($\beta = -1.097$, $SE = 0.347$, $p = 0.001$). As with the prior model, false alarms were not significantly predictive ($\beta = -0.031$, $SE = 0.035$, $p = 0.38$), nor was there an interaction between condition and experiment ($\beta = 0.511$, $SE = 0.435$, $p = 0.24$). Both exploratory models support the findings from Experiment 1 that both the pairing types and the participants own evaluation of them can be used to predict likelihood of re-representation.

A related question that arises from these data is: why are the analogy ratings for false analogies comparable to true analogies? The most direct explanation is simply that participants, acting in accord with the task demands, took any evidence of analogy as sufficient for proof of analogy. Recall that false analogies still shared higher-order relations between passages. If the ratings reflected participants’ confidence in their decision, the higher-order similarity may have been ample grounds for participants to rate the false analogies as more analogous than not. Despite this trend in responding, however, participants may not actually be processing these pairs the same way they would for true analogies.

4. General discussion

As a complete theory of analogy, SMT addresses the problem of flexibility by invoking the constraint of tiered identifiability and the mechanism of re-representation. The former filters out profligate matching (Gentner & Wolff, 2000) between possible relational predicates, while the latter allows for meaningful matches based on partial identity. Despite the importance of these theoretical components of SMT there exists relatively little supporting evidence. The amount of evidence in extant empirical work is disproportionately small given the critical impetus and scale of their proposed mechanisms difference. The work tends to consist of findings that are broadly consistent with the notion of re-representation rather than direct tests. More specifically, the few studies that do exist fail to show evidence of representational change specific to the relational predicates or to demonstrate a relationship between analogical similarity and representational change. The two experiments reported here were intended to conduct a more thorough and direct test of where the reasoner searches for a common superordinate relation (distinct from higher-order relations) that functionally satisfies partial identity—e.g., move being the superordinate of walk and run. Whereas minimal ascension describes a process of moving up a conceptual hierarchy, semantic decomposition posits that relations are decomposed into their constituent sub-elements. The partial identity is then attained by matches between these sub-elements. Differentiating between these two alternatives may be presently beyond the capabilities of behavioral methodologies.

Though re-representation is theoretically tied to structure-mapping, it is worth considering whether these data bear implications for any other models of analogy. As mentioned in the introduction, the LISA model of analogy does not entail re-representation, but does assume a representational format akin to semantic decomposition. In terms of how each model attains partial identity, the chief difference between LISA and SMT is that LISA always represents relations in terms of their constituent semantic primitives, while SMT only strategically re-presents relations in terms of their semantic primitives when necessary. The LISA model could potentially also account for the present findings if we assume that participants induce a separate schema from each of the original base-target pairings in the rating phase (Hummel & Holyoak, 2003). Consequently, the resulting abstractions would leave participants prone to the relational lures that are themselves abstractions. Though schemas have been shown to emerge from minimal comparison opportunities (Gick & Holyoak, 1983), it seems unlikely that observers would default to forming schemas from every comparison they encounter. With this in mind, however, the present experiments were not intended to provide delineating evidence between SME and LISA. Any hypothesis concerning schema abstractions should be tested directly in a study specifically designed to test that proposal.

4.2. Future directions

Apart from SMT and analogy more broadly, the present data also bear on related research in creativity and problem solving. For example, Dixon and Bangert (2002) investigated processes of theory re-description and revision as mechanisms of representational change in problem solving. They found that the manner and probability that people changed their representation of gear systems depended on the adequacy of the current representation for solving a given problem. While the impetus and scale of their proposed mechanisms differ from our current treatment of re-representation, it is not hard to imagine how re-representation at a larger scale might also facilitate changes to interpretations of whole propositions. Recent research (Oberholzer, Trench, Kurtz, & Minervino, 2018; Oberholzer, Trench, & Minervino, 2011) has revealed a possible mechanism by showing that analogical comparisons could prompt participants to re-represent the relational category (Gentner & Kurtz, 2005) of a target proposition to match the relational category of its paired base analog. Both re-representation of constituent propositional elements and of the relational categories of the whole propositions may be a route to creative processing. Dynamic re-construal of a representation at all levels is a promising potential avenue to overcoming functional fixedness and achieving convergent thinking.

Future work should also investigate possible educational applications. Analogy is a powerful tool for instruction (Richland & Simms, 2015). A better understanding of re-representation and its relation to analogy can help inform its use as a communicative device. Facilitating understanding of a subject can be achieved through either analogical retrieval or prompted evaluation, but the specific choice of which is used will have consequences for which base/target analogs are chosen and how they are implemented. If re-representation is a stronger prerequisite for analogical retrieval, relational predicates of target analogs should be expressed in domain-general forms as suggested by Clement, Mawby, and Giles (1994).
4.3. Conclusion

The present research uses a novel experimental approach to address a longstanding weak point in a leading theory of analogical processing. The findings are mostly consistent with the claims of SMT and go a considerable way to remediate the problematic lack of evidence for the key theoretical component of re-representation. Given the constraint of tiered identifiability, flexibility in analogy is attained via a representa- tional change of relational predicates (Yan et al., 2003). The current findings provide an empirical foundation for the theoretical claim of rerepresentation and adds to the impressive literature supporting the structure mapping theory of analogy.

5. Author notes

We would like to thank Drs. Celia Klin and Deanne Westerman, as well as the members of the LaRC Laboratory at Binghamton University for their helpful feedback.

Declaration of Competing Interest

None.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2019.04.031.

References