

# Investigating False Memory in the DRM Paradigm with Relational Category Content

Alexus S. Longo & Kenneth J. Kurtz

Department of Psychology, Binghamton University

Binghamton, NY 13905 USA

## Abstract

Fuzzy Trace Theory and the Activation Monitoring Framework disagree on whether gist or backwards associative strength best describes false memories in DRM. Recent evidence suggests situational features best describe processes underlying DRM results. Thematic, goal-derived, and relational categories capture different aspects of situational features, but perform differently in memory and coherence. We constructed novel category-specific lists for use in the DRM paradigm to determine whether different aspects of situational features make different contributions to a successful DRM result, whether relational and goal-derived content can produce false recognition despite low gist and backwards associative strength, and whether DRM captures an aspect of category coherence that has been difficult to measure in relational and goal-derived content. Only relational and thematic content produced sufficient false recognition. This provides mixed support for situational features, evidence that relational content makes specific contributions to DRM success, and evidence of coherence in relational categories.

**Keywords:** false memory; concepts and categories; relational cognition; category coherence; activation monitoring framework; fuzzy trace theory

## Introduction

Deese-Roediger-McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) classic false memory results have been replicated across a wide array of contexts, stimuli, and experimental manipulations. A successful DRM result requires that studied items and non-studied critical items (related to the studied content) are endorsed at recognition (or produced during recall) at comparable rates to each other, and at a significantly higher rate than for non-studied unrelated lures. DRM lists contain items sharing semantic similarity (i.e., taxonomic category, superficial and conceptual similarity), associative similarity (frequent co-occurrence), and orthographic or phonological similarity. The contributions of each type of similarity in producing DRM, and the underlying mechanisms of false memory production in DRM, remain under investigation (see Coane, McBride, Huff, Chang, Marsh, & Smith, 2021 for a review).

According to the Activation Monitoring Framework (AMF; Roediger, Watson, McDermott, & Gallo, 2001), activation spreads from activated (studied) items to

semantically or associatively connected non-studied items. If enough activation spreads to a non-studied item, a false memory is likely to occur (and error risk is reduced via monitoring). Based on the AMF, higher backwards associative strength (BAS)—or the strength of connections from studied items to the critical, non-studied item—predicts an increase in false memory (Roediger et al., 2001).

Fuzzy Trace Theory (FTT; see Reyna & Brainerd, 1995) claims that there are two memory traces. The *verbatim* trace is specific, decays rapidly, and acts to reduce the risk of memory error. The *gist* trace contains general associative information, slowly decays, and increases the risk of memory error. If a non-studied, critical item is sufficiently similar to the overarching theme of studied items, or the gist trace, then a false memory is likely to occur. According to FTT, a higher gist strength—how well the critical item represents studied items—predicts the likelihood of false memory.

There is mixed evidence for AMF and FTT. In support of AMF, BAS is a main predictor of false recall and recognition (Roediger et al., 2001), consistent BAS produces consistent levels of false memory even when varying the relatedness between critical and studied items (Hutchison & Balota, 2005), and a DRM result can still be found when BAS is high and gist strength is low (the critical item is not easily guessed from the studied items; Huff, Hutchison, Coane, Grasser, & Blais, 2012). Alternatively, other findings suggest that: BAS is only a better predictor of false recognition when gist strength is low (Brainerd, Chang, & Bialer, 2020), increased semantic similarity between studied and critical items increases false alarms when BAS is constant (Coane, McBride, Termonen, & Cutting, 2016), and BAS cannot always account for the different rates of false memory generated by different DRM stimuli (Brainerd, Yang, Reyna, Howe, & Mills, 2008; Cann, McRae, & Katz, 2011).

Mixed evidence is also found when analyzing characteristics of DRM stimuli. Roediger et al. (2001) performed a multiple regression analysis on factors related to false memory in DRM (length, concreteness, connectivity, true memory, BAS, forward associative strength, etc.). Roediger et al. (2001) found that true memory is negatively correlated to false memory while

BAS is positively correlated with false memory, supporting AMF. Brainerd et al., (2008) responded to these findings by performing a factor analysis on all DRM lists using 16 aspects of semantic content (e.g., concreteness, imagery, categorizability, synonymy, etc.). Brainerd et al. (2008)'s analyses show that veridical memory and false memory are associated with different factors—veridical memory is associated with imagery and concreteness, while false memory is associated with familiarity and meaningfulness—providing evidence for separate verbatim and gist traces.

Brainerd et al. (2008) further argues that BAS represents semantic processing as analyses found that false memory and mean BAS are both only associated with the familiarity/meaningfulness factor, thereby concluding that semantic processing, not BAS, explains DRM results.

The exact nature of the semantic processes involved is still under investigation. Brainerd et al. (2008)'s analyses included Wu and Barsalou's (2009) knowledge types. Knowledge types describe conceptual relations between items: synonym, antonym, taxonomic, situational, entity, and introspective. Each DRM list contains at least one knowledge type, and BAS positively correlates with the antonym, synonym, and taxonomic knowledge types (Brainerd et al., 2008). This furthers the argument that semantic processing instead of BAS best describes DRM results.

Cann et al. (2011) expand on Brainerd et al. (2008) by exploring how knowledge types, BAS, and gist strength contribute to false memories. Examining the knowledge types in DRM stimuli, situational, synonym, and taxonomic features predict both BAS and false recall. Situational features include roles, actions, locations, participants, and spatial relations—information typically used to describe scenarios and likely to produce strong gist. Cann et al. (2011) show that DRM-like lists with high situational features have high gist strength, low BAS, and perform similarly to high BAS DRM lists during recognition. While providing mixed support for AMF and FTT Cann et al. (2011) ultimately conclude that knowledge types may be a better basis for understanding DRM results, and provide a theoretical basis for BAS in AMF, and a potential operationalization of gist strength for FTT.

Cann et al. (2011) provide strong evidence for the importance of situational knowledge types—they explain variability in DRM list performance where BAS cannot, and gets around the issue of BAS and gist strength being highly related. We propose a further delineation of the situational knowledge type to further investigate specific semantic contributions to false memory production. The situational knowledge type includes action, function, location, participant, time, origin, manner, state of the world, etc. The categorization literature suggests that relational-role based similarity and thematic similarity are both represented in the situational knowledge type despite referring to very different types of similarity.

Relational-role categories contain items that play the same

role, i.e., *skin, jacket, lid* all play the role of a *cover*. In this way, the function and action aspects of the situational knowledge type describe relational-role based semantic similarity. Thematic categories contain items that frequently co-occur at the same time and place (desk, chalkboard, and window are frequently found in a classroom), similar to the time, location, and state of the world subdivisions of the situational knowledge type.

Not only do relational and thematic content capture different types of semantic similarity, they produce different amounts of gist. Thematic content greatly outperforms relational-role content when participants are asked to produce a label from a list of exemplars (Kurtz & Gentner, 2001). Thematic content also outperforms relational in general metrics of category coherence (Kurtz & Gentner, 2001) and in memory-based metrics of category coherence (Longo, 2024) as the lack of superficial feature overlap amongst members and lack of strict hierarchical structure within relational categories makes relational content difficult to assess and difficult to get participants to access. Cann et al. (2011) argue that the situational knowledge type may provide a basis for BAS and an index of gist strength—if the entirety of the situational knowledge type can account for this, then it would be expected that all aspects of the situational knowledge type contribute to BAS and gist. Differences in thematic and relational performance in the category literature however, provide evidence against this.

Expanding on Cann et al. (2011) to determine the most relevant aspects of the situational knowledge types to DRM is only one motivating factor in this research. From previous research, it is not likely that relational-role category content produces a strong gist trace (Kurtz & Gentner, 2001), so finding a DRM-like result in this content would be unexpected by FTT. Likewise, given the cross-cutting nature of relational content relative to taxonomic organization (see Gentner & Kurtz, 2005) it is unlikely that exemplars have strong associative connections to the category label. In the process of this work, the BAS of stimuli used was difficult to obtain via the free association norms available from Nelson, McEvoy, and Schreiber (2004). For all cue-target pairs in the norms for the relational critical items studied, the average BAS was below 0.01. That is, AMF would also predict poor DRM performance by relational stimuli. While this provides more reason to further investigate the situational knowledge types for more relevant distinctions/contributions, we are also interested in determining whether relational-role based category content makes an unexpected contribution to the semantic processes involved in DRM results—and if this contribution occurs despite the content being low in both gist strength and BAS.

The remaining motivation for this work comes from non-taxonomic category research. Unlike taxonomic categories, non-taxonomic categories do not require that members share broad feature overlap. Instead, items are members of the category if they match extrinsically, i.e., fulfilling a role or function (see Gentner & Kurtz, 2005).

Thematic, relational, and goal-derived categories are non-taxonomic. Goal-derived categories are much like relational categories to the effect that members play roles, but in goal-derived categories members play different roles to achieve some goal (e.g., various utensils are used to execute different aspects of the cooking process; Barsalou, 1983), while in relational categories all members play the same role. Such categories show coherence relative to ideals for a category rather than central tendencies or featural norms (Barsalou, 1985). The emphasis on extrinsic features results in non-taxonomic categories lacking semantic structure outside of adhering to a core relationship. This can make them difficult to study and could call into question their coherence and status as categories. Thematic categories are more well-defined in terms of co-occurrence in a particular spatiotemporal setting or situation. Despite lacking similarity among members, thematic categories remain high in category coherence. Relational and goal-derived perform poorly on metrics of category coherence – members within the same category are not rated very similarly, the category label is not easily guessed from a list of exemplars, and few and inconsistent exemplars are listed when participants are given a category label (Kurtz & Gentner, 2001). These findings have led to a kind of paradox—non-taxonomic content is important for reasoning, learning, and other higher order functions (Gentner & Asmuth, 2019; see Goldwater & Schalk, 2016), but relational and goal-derived content provide little behavioral evidence of their coherence or viability as categories. However, recent work using memory metrics to assess category coherence found that items within relational and goal-derived categories generate content-specific proactive interference and content relevant false recall (Longo, 2024), suggesting that relational and goal-derived content have an important, and testable, representation in memory.

Testing these stimuli in a DRM paradigm allows for this to be further investigated. The classic DRM result, through a category-lens, shows that items elicit (false) memory for their category label. If relational and goal-derived members can produce false recognition for the category label at rates comparable to classic DRM stimuli, then there would be more evidence for a solid connection between non-taxonomic category members and their overarching label. This would also show that one of the classic metrics of coherence – determining how well the category label is guessed from exemplars and vice versa – could be replaced with a DRM task to better capture category coherence in relational and goal-derived categories. According to the current state of category coherence investigations, thematic content is expected to generate a DRM result comparable to the classic stimuli, and relational and goal-derived content are not expected to produce a successful DRM result, but goal-derived content is likely to outperform relational.

We therefore test a combination of classic DRM stimuli, thematic content, and relational-role based content to assess whether further investigation into the unique contributions of aspects of the situational knowledge type to DRM results

is warranted, to determine if there is an important contribution from relational category content alone, and to further investigate the coherence for relational and goal-derived content. Goal-derived categories are also incorporated to further support our investigation into the importance of specific relational content. Specifically, does any potential contribution to a DRM result from relational stimuli come from the relations themselves or some overarching organizational process?

## Methods

### Subjects

53 SUNY Binghamton students participated for partial fulfillment of course requirements. 45 subjects were included in analysis (8 were removed;  $n = 1$  due to program error,  $n = 1$  for having higher false alarms than hits,  $n = 6$  for a hit rate of less than 50% on studied classic DRM stimuli).

### Stimuli

Four category types are tested (classic DRM, thematic, goal-derived, relational) with two sets of stimuli for each type. Classic DRM stimuli (Sleep, Needle) were taken from Roediger et al. (2001). Stimuli used for thematic, goal-derived, and relational content were built from word ratings collected in a previous experiment ( $n = 352$  subjects rated how well each item related to a concept; ratings range from positive to negative two). From this data, “items commonly found at the Beach/Movies” were selected as the thematic content<sup>1</sup> (average rating of 1.81), “items to pack for a Camping trip” and “things to do for weekend Entertainment” were selected as the goal-derived content<sup>2</sup> (average rating of 1.73), and “items that can be an Obstacle/Guide” were selected as the relational content<sup>3</sup> (average rating of 1.52). The distribution of rating scores for non-DRM stimuli used can be seen in Figure 1. Overall, there are eight lists of content as there are two lists for each category-type/condition. Only 14 words from each list were studied.<sup>4</sup> Unrelated lures were selected from unused content of each category type.

Since the non-DRM stimuli used were constructed from previous similarity rating data, we attempted to determine the BAS of our lists from Nelson et al., (2004), the bank of

---

<sup>1</sup> Beach included *sand, surfboard, lifeguard*; Movie included *film, screen, popcorn*.

<sup>2</sup> Entertainment contained *shopping, festival, bowling*; Camp contained *tent, firewood, repellant*.

<sup>3</sup> Obstacle included *hurdle, problem, wall*; Guide included *teacher, handbook, navigator*.

<sup>4</sup> One word was removed from each DRM list. “Knitting” was removed from “Needle” as it can be entertainment (goal-derived content); removing it increases the mean BAS “Needle” to 0.208. “Blanket” was removed from “Sleep” as it could be a cover (unrelated lure category); removing it increases the mean BAS of “Sleep” to 0.460. These removals should not impact the data (AMF suggests these removals lead to a stronger DRM result).

norms from which the BAS for DRM stimuli can be found. Unfortunately, many of the studied words used do not appear in the database. For our thematic content, the average BAS is .320 (from ten items), the average BAS of goal-derived content is .092 (five items), and the average BAS of the six relational items found is .004. To provide an idea about the general BAS of the content used, we averaged the BAS for all target items (whether used in our stimuli or not) for each cue (critical item). The same pattern in BAS holds (thematic = .154; goal-derived = .017; relational = .009). This, in addition to the spread of word similarity rating data (Figure 1), suggest that the relational content used has the lowest BAS.

Lastly, differences in letter length were kept minimal. DRM stimuli has an average letter length of 5.14, Thematic 6.86, Goal-Derived 7.22, and Relational has an average length of 8.07. Usage frequency (from Brysbaert & New, 2009) varied given the nature of the stimuli. DRM content has an average usage frequency of 55.16, Thematic 11.52, Goal-Derived 13.42, and Relational 23.22.

## Procedure

Before participating, subjects were informed of the design (blocked study-distractor phases followed by a total intermixed recognition test). For the recognition test, participants were told to select “YES” if they remembered studying that item at any point during the experiment. They were told the test will contain studied and non-studied items, and that they should not guess.

The order of lists studied, and the order of content within each list, was randomized. During each study phase, 14 items from one list were shown individually for 2s, interspersed by a 0.5s fixation cross. After each study phase

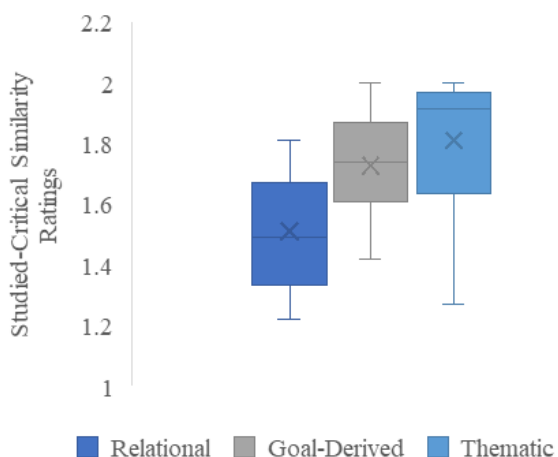


Figure 1: Box and whisker plots depicting the spread of similarity ratings for the relational (dark blue), goal-derived (gray), and thematic (light blue) content used. The ‘X’ indicates the mean value.

there was a visual arithmetic distractor task for 30s. After all eight study-distractor blocks, participants studied a total of 112 items. All participants studied both sets of stimuli for each category type of interest.

The recognition test was self-paced. Items were randomly presented in an intermixed order. For each item participants were asked if the word shown was one they studied, and they are reminded to not guess. Participants had the option of selecting a YES or NO button. During recognition seven studied items per list (56 total studied items), one critical lure per list (eight total), and 12 unrelated lures per category-type (48 non-studied, non-critical items) were displayed. Overall, 112 items were shown during the recognition test (half studied).

## Design

We used a blocked study-distractor before intermixed recognition test design based on Coane et al., (2016). All participants experienced all lists, so all factors are within-subjects. All data for each category type are combined, and then a one-way repeated measures ANOVA is used to compare the proportion of ‘yes’ responses at recognition between studied, non-studied, and critical items for each category type (that is, for example, both DRM list data are combined and one ANOVA is performed on that data). Follow up t-tests are performed when necessary (p-values are Bonferroni corrected for three comparisons). A successful DRM-like result has three requirements: (1) Studied items have a significantly higher proportion of ‘yes’ responses than non-studied items; (2) Critical items have a significantly higher proportion of ‘yes’ responses than non-studied items; (3) Proportion of ‘yes’ responses for studied and critical items is not significantly different.

Previously, the potency of a DRM list is measured by the proportion of false alarms generated (Roediger et al., 2001). Since memory for goal-derived and relational content is worse than for other types of content (Asmuth & Gentner, 2005; Barsalou, 1992, 1993; Gentner, Rattermann, & Forbus, 1993) ranking the amount of false alarms produced may artificially disadvantage relational and goal-derived content compared to other content-types. To resolve this, a difference score is taken from each participant (proportion of ‘yes’ responses to studied items subtracted by the proportion of ‘yes’ responses to critical items) for each category type tested. A positive difference score indicates more hits than false alarms, while a negative difference score indicates more false alarms than hits. A one-sample t-test is used to compare the average difference score of each category type to an ideal DRM result – in which the difference between studied items and critical items is zero.

## Results

### Evaluating DRM Success

For each condition/content-type, data from the two relevant lists were combined before analyses. The first analysis for each condition is a one-way repeated measures ANOVA

Table 1: Descriptive statistics (standard deviations in parentheses) and results of the repeated measures one-way ANOVA are reported for the proportion of ‘yes’ responses at recognition for each item type within each condition.

	Descriptive Statistics			F-Test		
	Studied	Non-Studied	Critical	F-value	p-value	$n^2$
DRM	.76 (.15)	.13 (.15)	.76 (.36)	$F(2, 42) = 118.18$	< .001	0.73
Thematic	.75 (.18)	.13 (.15)	.71 (.38)	$F(2, 42) = 91.25$	< .001	0.68
Goal-Derived	.62 (.21)	.13 (.15)	.48 (.37)	$F(2, 42) = 56.69$	< .001	0.56
Relational	.67 (.20)	.13 (.15)	.57 (.38)	$F(2, 42) = 70.39$	< .001	0.62

comparing the proportion of ‘yes’ responses across item-type (studied, non-studied/lure, critical). All values were significant (see Table 1).

Our classic DRM lists successfully replicate classic DRM results. Recognition endorsement of studied items was significantly higher than for non-studied items,  $t(44) = 13.37, p_B < .001, d = 2.61$ , but not significantly different from critical items,  $t(44) = .11, p_B = 1.00$ . Critical item endorsement was significantly higher than non-studied item endorsement,  $t(44) = 13.26, p_B < .001, d = 2.59$ .

There is also a DRM result for thematic content. Endorsement for critical items and studied items did not significantly differ,  $t(44) = .82, p_B = 1.00$ , while both critical items,  $t(44) = 11.27, p_B < .001, d = 2.28$ , and studied items,  $t(44) = 12.09, p_B < .001, d = 2.44$ , received significantly more ‘yes’ responses during recognition than non-studied.

Likewise, relational content produced a successful DRM result. More studied items were endorsed than non-studied items,  $t(44) = 11.16, p_B < .001, d = 2.06$ , more critical items were endorsed than non-studied items,  $t(44) = 9.07, p_B < .001, d = 1.67$ , and there was no significant difference between studied and critical items,  $t(44) = 2.10, p_B = .12$ .

The goal-derived condition did not meet all requirements for a successful DRM result. Studied  $t(44) = 10.36, p_B < .001, d = 1.91$ , and critical items  $t(44) = 7.30, p_B < .001, d = 1.34$ , were endorsed significantly more than non-studied, but studied and critical items also significantly differed,  $t(44) = 3.07, p_B < .01, d = .57$ . While critical lures were

more difficult to reject at recognition than unrelated content, they were still more easily rejected than studied content.

### Comparing Difference Scores

Within each condition, difference scores are calculated by subtracting the proportion of ‘yes’ responses to critical items from that of studied items. Difference scores close to zero indicate no difference in recognition endorsement between studied and critical items; positive scores indicate more studied item endorsement. A one-sample t-test is used to compare the difference scores of each condition to an ideal value of zero. All difference score descriptives and t-test values are reported in Table 2. Relational content approached significance ( $p = .06, d = .29$ ), but goal-derived content was the only significant value,  $t(44) = 2.70, p < .05, d = .40$ . Classic DRM, thematic, and relational stimuli produced a classic DRM result; recognition endorsement did not significantly differ between studied and critical items.

### Discussion

We set out to determine whether specific aspects of the situational knowledge type should be further investigated for independent contribution to DRM results. We also wanted to determine whether relational role-based content makes a specific contribution to DRM-relevant semantic processes despite having low gist strength and BAS. Further, we sought to determine if this contribution was due to pure relational similarity or if overarching contextual similarity (with role based differences) would produce a stronger DRM result. Lastly, we sought to explore whether testing relational and goal-derived content in a DRM paradigm would provide support for their coherence.

Thematic content and relational content were split into different lists as this content represents different aspects of the situational knowledge type. Thematic pertains to time, location, and state of the world, while relational content refers more to actors and functions. A successful DRM result was found in both types of content, supporting the conclusions of Cann et al. (2011); the situational knowledge type makes some contribution to false recognition in DRM. Our results suggest that the situational knowledge type does not require further parsing-that both relational and thematic content within this knowledge type make contributions to successful DRM results.

Table 2: Descriptive statistics for the difference scores between studied and critical item endorsement, and results of the one-sample t-test used to compare the score in each condition to an ideal value of zero.

	Difference Scores		One-Sample t-Test	
	Mean	SD	t-value	p-value
	DRM	.01	.37	$t(44) = 0.09$
Thematic	.04	.42	$t(44) = 0.68$	.50
Goal-Derived	.15	.36	$t(44) = 2.70$	.01
Relational	.10	.34	$t(44) = 1.97$	.06

Additionally, Cann et al., (2011) found that false recognition in high situational knowledge type lists with high gist strength and low BAS occurred at comparable rates to high BAS DRM lists. Our relational stimuli lack both gist and BAS, but not situational features, providing stronger support for Cann et al. (2011)'s argument that situational knowledge types may better explain DRM results. However, goal-oriented content should contain comparable levels of situational features to relational content, but did not produce as strong of a successful DRM result. High situational features may not always elicit high levels of false recognition.

Relational content was contrasted with goal-oriented content to determine whether similarity in roles or context produces a successful DRM result. Data suggest that pure relational-role similarity is important for generating a DRM result. Since this content should be similarly high in situational features, pure relational-role similarity may make contributions independent to situational features. Coane et al., (2016) found what they called a category boost effect for DRM lists constructed to contain taxonomic category similarity. While relational and taxonomic categories do not often perform similarly (Kurtz & Gentner, 2001; Longo, 2024), we provide preliminary evidence for a similar relational similarity based boost effect. However, more investigation on this is necessary, especially considering that the goal-oriented content did get close to a successful DRM result.

The successful DRM result found in relational content (and somewhat in goal-derived content) despite suspected low BAS and gist strength is of particular interest. Since our relational content has low BAS, this result contradicts predictions of AMF and adds to increasing evidence that semantic processing underlies false memory in DRM (Brainerd et al., 2008; Cann et al., 2011). Since relational content produces behavioral results indicating low gist strength (Kurtz & Gentner, 2001), this finding also contradicts predictions of FTT. It is not likely that other aspects of the stimuli used can account for the results either. Average letter length did not greatly differ between the stimuli used, and numerically, relational content had the longest average length. Usage frequency (from Brysbaert & New, 2009) did vary between content, but thematic produced a strong DRM result with the lowest average frequency. While relational content had the second highest usage frequency (23.22), it is still numerically below that of the DRM content used (55.16) and closer to that of thematic (11.52) and goal-oriented (13.42). So, usage frequency differences cannot explain our relational DRM result. However, goal-oriented content having the second lowest frequency may explain why it produced only a near-successful DRM result. When referring to the Nelson Norms (Nelson et al., 2004) for the forward associative strength (FAS) of stimuli used, relational content has the lowest average FAS while DRM has the highest.<sup>5</sup> This

---

<sup>5</sup>  $FAS_{DRM} = 0.113$  (fifteen items);  $FAS_{Thematic} = 0.099$  (ten items);  $FAS_{Goal-Derived} = 0.057$  (five items);  $FAS_{Relational} = 0.038$  (ten items).

pattern holds when averaging FAS for all items listed with our critical words in the Nelson Norms.<sup>6</sup> Neither length, usage frequency, FAS, BAS, or gist strength account for the successful relational DRM result. This provides further evidence for a relational similarity based boost effect and that current measures of category coherence do not accurately capture the connectedness of relational and goal-oriented content.

Metrics of category coherence show that relational and goal-derived perform similarly to each other and far worse than thematic and taxonomic (Kurtz & Gentner, 2001; Longo, 2024). This leads to a sort of paradox: Relational and goal-oriented content are used in problem solving and other aspects of higher order cognition (Gentner & Asmuth, 2019; see Goldwater & Schalk, 2016), but can relational and goal-derived categories be meaningful collections of items if members are not similar to each other, the label is not apparent from its members, and its members are not easily accessed from the label? Our successful DRM result suggests that relational role-based categories do point to interconnected members in a meaningful category-specific way. This work argues that memory metrics, such as DRM, may be useful measures of category coherence, leading to a new way to assess how non-taxonomic and non-associative content is organized and accessed.

Additionally, the performance difference observed between relational and goal-derived content is not typically seen. When there are coherence differences in relational and goal-derived content, goal-derived content outperforms relational (Kurtz & Gentner, 2001; Longo, 2024). This is attributed to the contextual or thematic content present in goal-derived categories that is not present in relational. However, our data suggest that pure relational-role similarity may be more important, in terms of DRM, than differences in roles that maintain similarity in context. Our data also suggest that relational content may have stronger connections between members and the overarching category label than previously reported (Kurtz & Gentner, 2001).

Overall, these findings support Cann et al., (2011). The situational knowledge type may explain successful DRM when there is low gist strength and low BAS, and in our case, where neither AMF nor FTT predict the successful relational DRM result. Cann et al. (2011) argue that the situational knowledge type underlies BAS and may outline a defining basis for gist strength. We have shown that content lacking in BAS and gist strength, but built from specific aspects of the situational knowledge type, produces a successful DRM result, and this may only occur if there is also pure relational role based similarity. This supports further investigation into situational knowledge types, and the unique contribution of relational role based similarity in the semantic processes underlying DRM. These findings also support further exploration of alternative methods to measuring category coherence in relational and goal-derived categories.

---

<sup>6</sup>  $FAS_{DRM} = 0.097$ ;  $FAS_{Thematic} = 0.048$ ;  $FAS_{Goal-Derived} = 0.049$ ;  $FAS_{Relational} = 0.042$ .

## References

- Asmuth, J. A., & Gentner, D. (2005). Context sensitivity of relational nouns, *Proceedings of the Twenty-seventh Annual Meeting of the Cognitive Science Society*, (pp. 163-168).
- Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts, *Memory & Cognition*, *10*(1), 82-93.
- Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(4), 629-654.
- Brainerd, C. J., Yang, Y., Reyna, V. F., Howe, M. L., & Mills, B. A. (2008). Semantic processing in "associative" false memory, *Psychonomic Bulletin & Review*, *15*(6), 1035-1053.
- Brainerd, C. J., Chang, M., & Bialer, D. M. (2020). From association to gist. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(11), 2106.
- Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American-English, *Behavior Research Methods*, *41*(4), 977-990.
- Coane, J. H., McBride, D. M., Huff, M. J., Chang, K., Marsh, E. M., & Smith, K. A. (2021). Manipulations of list type in the DRM paradigm: A review of how structural and conceptual similarity affect false memory, *Frontiers in Psychology*, *12*.
- Coane, J. H., McBride, D. M., Termonen, M.-L., & Cutting, J. C. (2016). Categorical and associative relations increase false memory relative to purely associative relations, *Memory & Cognition*, *44*(1), 37-49.
- Cann, D., McRae, K., & Katz, A. (2011). False recall in the Deese-Roediger-McDermott paradigm: The roles of gist and associative strength, *Quarterly Journal of Experimental Psychology*, *64*(8), 1515-1542.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall, *Journal of Experimental Psychology*, *58*(1), 17-22.
- Gentner, D., & Asmuth, J. (2019). Metaphoric extension, relational categories, and abstraction, *Language, Cognition, and Neuroscience*, *34*(10), 1298-1307.
- Gentner, D., & Kurtz, K. (2005). Learning and using relational categories. In W.K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman & P. W. Wolff (Eds.), *Categorization inside and outside the laboratory: Festschrift in honor of Douglas L. Medin*. Washington, DC: APA.
- Gentner, D., Rattermann, M. J., & Forbus, K. D. (1993). The roles of similarity in transfer: Separating retrievability from inferential soundness, *Cognitive Psychology*, *25*(4), 524-575.
- Goldwater, M., & Schalk, L. (2016). Relational categories as a bridge between cognitive and educational research, *Psychological Bulletin*, *142*(7), 729-757.
- Huff, M. J., Hutchison, K. A., Coane, J. H., Grasser, E. B., & Blais, J. E. (2012). Interpolated task effects on direct and mediated false recognition: Effects of initial recall, recognition, and the ironic effect of guessing, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1720-1730.
- Hutchison, K. A., & Balota, D. A. (2005). Decoupling semantic and associative information in false memories: Explorations with semantically ambiguous and unambiguous critical lures, *Journal of Memory and Language*, *52*(1).
- Kurtz, K. J., & Gentner, D. (2001). Kinds of kinds: Sources of category coherence. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, (pp. 522-527). Mahwah, NJ: Lawrence Erlbaum Associates.
- Longo, A. (2024). Memory and coherence metrics in non-taxonomic categories: Release from proactive interference in thematic, goal-derived, and relational content. Master's thesis, Psychology Department, State University of New York Binghamton. ProQuest.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms, *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402-407.
- Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis, *Learning and Individual Differences*, *7*(1).
- Roediger, H., & McDermott, K. (1995). Creating false memories: Remembering words not presented in lists, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803-814.
- Roediger, H., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis, *Psychonomic Bulletin Review*, *8*(3), 385-407.
- Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation, *Acta Psychologica*, *132*, 173-189.