A Framework for Predicting Memory Errors with a Bayesian Model of Concept Generalization

Isabella C. DeStefano (idestefa@ucsd.edu),
Timothy F. Brady (tfbrady@ucsd.edu), Edward Vul (evul@ucsd.edu)
University of California, San Diego, Department of Psychology
9500 Gilman Dr., La Jolla, CA 92093 USA

Abstract

"Similarity" is often thought to dictate memory errors. For example, in visual memory, memory judgements of lures are related to their psychophysical similarity to targets: an approximately exponential function in stimulus space (Schurgin et al. 2020). However, similarity is ill-defined for more complex stimuli, and memory errors seem to depend on all the remembered items, not just pairwise similarity. Such effects can be captured by a model that views similarity as a byproduct of Bayesian generalization (Tenenbaum & Griffiths, 2001). Here we ask whether the propensity of people to generalize from a set to an item predicts memory errors to that item. We use the "number game" generalization task to collect human judgements about set membership for symbolic numbers and show that memory errors for numbers are consistent with these generalization judgements rather than pairwise similarity. These results suggest that generalization propensity, rather than "similarity", drives memory errors.

Keywords: Concepts and Categories, Memory, Generalization, Similarity

Introduction

One of memory's defining properties is its fallibility and tendency for systematic error (e.g., Bartlett, 1932). Thus, empirical measures of memory often focus on memory errors and the extent to which such errors are made to similar vs. dissimilar items. For example, in low-level domains like visual working memory, the pattern of errors people make to similar items is often termed memory "precision" and models of its properties are leveraged to understand memory performance and capacity (e.g., van den Berg et al. 2012). Likewise, in long-term recognition memory, the mistaken recognition of an item that is similar, but not identical, to a previously encountered item is a ubiquitous and robust memory error, referred to as gist-based false recognition (Koutstaal & Schacter, 1997). One particularly popular design to demonstrate this is the DRM paradigm (Roediger & McDermott, 1995; Gallo, 2010), which is used to study false memory. In the DRM, false memories arise when a subject is asked to remember a list of associated words (e.g., bagel, eggs, bacon) then, when given a memory test, they later recall or recognize related words that were not originally on the list (e.g., breakfast). The DRM is critically dependent on the relatedness or similarity of the misremembered word to the encoded exemplars.

Formal models have attempted to explain why we make such memory errors by connecting "similarity" to memory. For example, exemplar-based models of visual recognition propose particular functional forms for similarity, and link such similarity functions not only to memory but also to categorization and generalization (Nosofsky, 1992). Likewise, spreading activation models postulate that remembered words activate semantically similar items, which then are more likely to be falsely recalled (Roediger & McDermott, 2000). Even rich patterns of errors can be thought of in this way: For example, Schurgin et. al. (2020) show that in the domain of visual memory — particularly continuous reproduction of simple visual features, like color hue — signal detection decisions based on item similarities can explain many memory errors. In these simple metric spaces, pairwise similarity is well approximated by an exponential function of distance in the underlying stimulus space (Shepard, 1987), so Schurgin et al. link this similarity function to an activation for each item and, via signal detection, predict entire patterns of participants' memory errors based on these activations.

Although these examples suggest that some notion of "similarity" underlies memory errors, pairwise similarity has little explanatory force on its own — and breaks down when we consider either high-level domains without well-defined, metric perceptual spaces, or presentation of entire sets of items. What defines the similarity between two words, or two more abstract concepts? Although such similarities may be empirically measured, they are much harder to predict based on distances in a latent space (Mikolov et al., 2013) and other, non-metric conceptual structures may be needed (Shepard, 1980). Furthermore, the rated similarity between two items varies as a function of context (Tversky 1977), suggesting that the pairwise similarity is not a stable relational property. Finally, when considering how items are related to a whole set, pairwise similarity is inadequate: although there have been many attempts to sum (Nosofsky 1986), average (Ashby & Leola Reese 1995), or maximize (Goldstone 1994) pairwise similarities to obtain a single score for whole set similarity, no stable rule seems to capture human behavior.

Tenenbaum & Griffiths (2001) propose to resolve the challenges with defining similarity by suggesting that generalization, rather than similarity, is the core concept.

That is, when we see a set of items, we consider what compressed conceptual representations are consistent with that set. When asked how similar a given target is to that set, we answer the question "how likely is that target item to be an exemplar of the concepts suggested by the set?" On this account, similarity is epiphenomenal to the more general process of conceptual generalization, and when asked for our subjective similarity we provide ratings by simply indicating our generalization propensity. This account removes the reliance on pairwise similarity, and substitutes for it the notion of set-to-item generalization. Consequently, this formulation applies unaltered to abstract categories without metric latent spaces, and to sets comprised of multiple items. Moreover, this account explains a number of puzzles about similarity ratings: why they are often asymmetric, violate the triangle inequality, and vary with context (e.g., Tenenbaum & Griffiths, 2001). Finally, this account also has the appealing property of reducing to known rules about pairwise similarity (e.g., Shepard, 1987) in the limiting case where the set includes just one item.

In the current paper, we seek to establish a framework for linking generalization directly to memory errors. In particular, we suggest that combining a signal detection-based memory model, like that of Schurgin et al. (2020), with a Bayesian framework for generalization, like that proposed by Tenenbaum and Griffiths (2001), can bridge the gap between models of similarity-based memory errors and higher-level tasks with richer, more contextual errors, and provide both precise prediction of memory errors as well as significant theoretical clarity above and beyond notions of "similarity".

To test this, we ask if the propensity to generalize from a set to a given stimulus underlies the likelihood of false alarming to that stimulus in a memory task. We focus on the domain of remembering symbolic numbers, where generalization is much richer than for simple visual stimuli like colors, but can nevertheless be formalized (e.g., Tenenbaum & Griffiths, 2001). In this domain, the Bayesian framework for generalization naturally subsumes two classes of generalization: that based on magnitude (which account for exemplar-based "similarity" functions) and rule-based generalizations (Tenenbaum & Griffiths, 2001). Then, by considering the memory co-activations implied by this Bayesian model of generalization, and applying a memory model on top of them (Schurgin et al. 2020), we propose that it may be possible to predict memory errors in a way that cannot be done from pairwise similarity data alone. The key predictions of our framework are that:

(a) Propensity to generalize between items, even when such propensity does not follow simple rules of similarity (like in a metric space), will predict memory errors.
(b) Multiple observations (e.g. multiple items stored in memory) jointly determine this generalization propensity and thus these memory errors.

In the following we outline such a framework and suggest theoretical connections between models of generalization and memory and highlight what formalisms would be important to include in a unifying model. We then introduce a novel paradigm with two experiments based on the 'number game' that bridges the gap between research on generalization and research on memory errors. In the first experiment, we elicited generalization data from participants in response to sets of numbers, asking them to generate other numbers that would fit in that same set. Then in a second experiment, we used this data to predict memory performance. We gave people the same sets of numbers to remember, and probed their memory in a 3-alternative forced-choice task. We found that the rate at which a particular number elicits memory errors for a given set is related to the propensity of that number to be generalized from that set in Experiment 1.

A Framework for Generalization-based Memory Errors

A framework unifying metric and conceptual models of memory errors, must rely on two components: (1) A model of item co-activation that does not rely on pairwise similarity between items, but instead relies on generalization propensity and (2) a model that can predict human memory performance on a variety of tasks given the item activations.

Concept Generalization as the Basis of Memory activation. To formalize item co-activation in a manner that can apply equally well to metric and more general conceptual spaces, we turn to a Bayesian framework for concept generalization put forth by Tenenbaum (2000) and later extended by Tenenbaum and Griffths (2001). This account unifies two kinds of generalization that were previously thought to be distinct processes: abstracting rules and generalizing based on exemplars. On this account, the probability that a new object y is an element of some compressed representation (or concept) C inferred from exemplars $X = \{x^{(1)}, \ldots, x^{(n)}\}$ can be calculated by marginalizing over concepts likely to describe X:

$$p(y \in C \mid X) = \sum_{c \in C} p(y \in c) p(c \mid X)$$

This formalism has the advantage of being applicable to all conceptual spaces, metric or otherwise, and is consistent with extracting compressed, more abstract representations of items in terms of the hypotheses they are consistent with (Brady & Tenenbaum 2013).

Tenenbaum (2000) first used this formulation in the domain of symbolic numbers, because numerical reasoning seems to use both rule- and exemplar-based generalization. Their model predicted data from an empirical generalization task using i) a single exemplar, ii) a set of exemplars constrained by a simple conceptual rule, and iii) sets of exemplars that were of similar magnitude (Tenenbaum, 2000). In particular, the task they worked with — 'the number game' — involved a 'computer' that spit out a set of numbers, after which participants had to indicate which other numbers the computer would likely accept (i.e., they had to generalize from the given examples).

To model this task, Tenenbaum (2000) assumed that people had a structured set of prior knowledge about numbers that formed their hypothesis space for generalization. These hypotheses included ranges (e.g., numbers 15-30), as well as rules (e.g., multiples of 10). Their likelihood function was based on the 'size principle' reflecting that hypotheses that included fewer numbers were more likely to generate any one of those numbers.

This model captured basic patterns in human data better than competing models by unifying rule-based and exemplar-based generalization. When all the weight ends up on a single hypothesis (e.g., powers of 10), the model predicts rule-like generalization; however, when many hypotheses are in play (e.g., as when a single number is the basis of generalization), the model marginalizes over many plausible intervals and yields graded, exponential-like fall-off of generalization, as in models of pairwise generalization and similarity (e.g., Nosofsky, 1992; Shepard, 1987).

We take this model as our basis for understanding how people will generalize from numbers; and thus, how a set of observed numbers will yield memory activations.

Predicting Memory from Generalization Propensities

If you are holding in mind a number — or a set of numbers — intuition suggests you will be more likely to falsely remember a similar number rather than a dissimilar number. A recent model (TCC, Schurgin et al. 2020) suggests that such "similarity"-based errors can be formalized using the knowledge that when an item is encoded, this causes activation or familiarity not only for that item, but also for other items in proportion to their psychological similarity to the target item. After such activations are corrupted by noise (i.e. according to signal detection theory), they serve as the decision variable people use to choose which items have previously been seen.

In the simple perceptual spaces that model was applied to (Schurgin et al. 2020), the amount of activation that a non-encoded item receives was an approximately exponential function of its distance from the encoded item in psychological space, consistent with pairwise notions of similarity (e.g., Nosofsky, 1992) and the universal law of generalization (Shepard, 1987). In particular, in the domain of color, if you see a red item, you get a large boost of activation for red; a small boost for orange; and effectively no boost at all for green or blue, which are both far enough away from red to be approximately maximally dissimilar. All activations are then corrupted by noise, and when faced with a choice between multiple colors in a memory probe, participants report the largest activation as the 'old' item.

Formally, if item t is encoded, the mean memory-match signal for a given item x on the working memory task is given by $d_x = d' f_t(x)$, where d' corresponds to memory strength and $f_t(x)$ is a function describing the pairwise similarity of each item to the encoded item t. When x = t, $f_t(x) = 1$, so $d_0 = d'$. Then, as noted above, during the memory probe test each item that is shown as a possible response option generates a memory-match signal, m_x, conceptualized as a random draw from that item's memory strength distribution, which is centered on d_x but, consistent with signal detection theory, is corrupted by noise. That is, $m_x \sim N(d_x, 1)$. The response, r, on a given trial is made to the item that generates the maximum memory-match signal, $r = argmax(m)$.

Rather than conceiving of the memory activation function $f(x)$ as following a simple pairwise similarity that scales exponentially as a function of pairwise item distance, here we propose this activation should instead rely on the posterior predictive distribution over all possible generalizations (integers between 0 and 100) derived from human generalization data or the Griffiths and Tenenbaum (2001) model. This replaces the concept of pairwise similarity as the cause of memory activation with the more general idea of generalization propensity as the cause of such activations.

Griffiths and Tenenbaum (2001), following Shepard (1987), showed that with single examples that are largely captured by magnitude-based hypotheses (e.g., interval set ranging from 50 to 60), the exponentially decreasing function of distance that is generally observed in simple stimuli (e.g., in TCC) is naturally generated by their model. This is because the Bayesian model weights all possible intervals that include that item from the prior and weighs the smaller ones more heavily than the larger ones (according to the size principle) — resulting in an approximately exponential fall-off of generalization likelihood. Thus, the Bayesian model of generalization as applied to memory is also a true generalization of TCC, as it includes TCC as a special case. In particular, the predictions made about generalization from a single exemplar will approximate the exponential law of generalization used in Schurgin et al. (2020) as a model of similarity in metric psychological spaces (like the color space used in visual memory).

The generalization function for rule-based compression will not approximate an exponential, however. Given examples like 10, 20, 30 and 40, people will be unlikely to false alarm to 23; instead, the posterior predictive over all possible compressions (both interval- and rule-based), will instead represent a more rule-like, gist-based representation, rather than an exponential distance from exemplars. However, the generalized TCC framework still suggests that activation in memory will be proportional to the activation via the generalization function, such that numbers that are more likely to arise in generalization by definition also enjoy higher memory co-activations — even if these numbers are far from the seen numbers (e.g., 80).

Summary of Framework We propose that combining the structured generalization framework of Tenenbaum and Griffiths (2001) with the TCC framework for activation in memory and subsequent memory decisions (Schurgin et al. 2020), should allow us to predict memory performance and errors from sets of abstract items. In particular, this framework provides a way of thinking about false memories of related items that arise in a variety of tasks (like the DRM; Gallo, 2010), but with a formal framework that defines exactly which items are more or less related to a presented

Table 1: Classes of sets used to generate stimuli.

Set Type	Description	Set Constraints	N
Interval	8 sequential integers	$x \in [a,b]\ ;\ \lvert[a,b]\rvert = 8$	10
Interval	16 sequential integers	$x \in [a,b]\ ;\ \lvert[a,b]\rvert = 16$	10
Interval	32 sequential integers	$x \in [a,b]\ ;\ \lvert[a,b]\rvert = 32$	10
Rule	Multiples of a	$x \bmod a \equiv 0\ ;\ a \in [2,12]$	12
Rule	Numbers ending in b	$x \bmod 10 \equiv b\ ;\ b \in [1,9]$	10
Rule	Non-intuitive symbolic rules	$x \bmod a \equiv b\ \ b \in [1,9]\ ;\ a \in [2,12]$	8

set, and how such relations feed into memory decision processes.

To provide preliminary evidence for this framework, we used specific sets of 4 numbers to elicit generalizations in Experiment 1, then used that generalization data to design the stimuli and memory probes in Experiment 2. The sets of 4 numbers sometimes inspired rule-based generalization, sometimes exemplar-based generalization, and sometimes were unintuitive according to either rules or exemplars, to provide a test of whether generalization would predict memory even when generalization was not itself straightforward to model. These data provided empirical evidence about the structure of number space and the activations of the items after seeing each of the sets.

Experiment 1: Generalization Task

Methods

Participants. A total of 198 subjects provided responses. Subjects were excluded from the final set of data based on their performance on the interval sets (excluded if more than half of their responses were not within the interval), resulting in a total of 171 subjects included in the final data set.

Stimuli and Procedure. There were 60 stimuli sets in total. Stimuli sets were generated by quasi-randomly sampling four numbers from the positive integers between 0 and 100 that belonged to one of 6 set classes described in Table 1. The ten rule-based sets were sampled from each class of type interval (8-span, 16-span, and 32-span). All 9 multiple and all 11 digit-1 classes were used to generate stimuli. The non-intuitive classes of sets follow the same symbolic form as the multiple and digit-1 sets, but are more difficult to describe verbally (for example "a number whose remainder when divided by 8 is equal to 6"). The stimuli used to represent each set were 4 numbers sampled from the set using a normalized exponential distribution, such that the numbers in the set with smaller magnitude were more likely to be chosen as exemplars Each subject completed 60 trials, one for each of the stimuli sets. The order of presentation was randomized across subjects. For each set, four specific exemplars were presented that were identical for each subject. The order in which the exemplars appeared on each trial was randomized. On each trial, subjects were shown 4 exemplar numbers simultaneously and were asked to generate another number from this set, and gave two unique responses before receiving feedback about how many of their responses on that trial were in the predetermined set. A response was considered correct if it belonged to the set from which the exemplar stimuli were sampled. A running tally of their correct responses were also displayed as an overall score for feedback.

Results and Discussion

The data collected in Experiment 1 was broadly consistent with the patterns observed in Tenenbaum (2000). Example sets of responses collected from four of the stimuli sets are shown in Figure 1. The interval sets followed the expected exponential decrease in generalization as a function of distance from an exemplar (Fig. 1A; Shepard, 1987).

The rule-based sets of class "multiples" showed the expected all-or-none pattern of generalization, with very few responses failing to meet the class rule (Fig 1B). An interesting pattern of results here that was not observed in the data presented by Tenenbaum (2000) is that while the reports follow the rule, there appears to be a distance effect as well: there is an approximately exponential decrease in reports within the category as a function of distance from the center of the shown examples. This is likely a consequence of our task asking participants to generate only 2 samples, rather than rate how good of a fit a variety of numbers were (as in Tenenbaum, 2000).

The "non-intuitive" class sets had unexpected qualitative patterns. While there appear to be consistent patterns of generalization for these sets, they do not necessarily reflect the pattern implied by the non-intuitive rule. Some of these consistencies may come from other rule-based hypotheses. For example, in the data shown in Fig. 1C the most frequent response followed the non-intuitive rule, but many other responses did not. This is likely because participants inferred the more intuitive rule "multiples of 3" instead of inferring that the sequence of numbers incrementally increased by 6.

The consistent generalization patterns may also come from generalizing from subsets of the exemplars. For example, in Figure 1D many frequent responses followed the rule x^2, when this rule applied to only three of the exemplars. This suggests that generalizations could be made from a subset of exemplars or that outlier exemplars are discounted

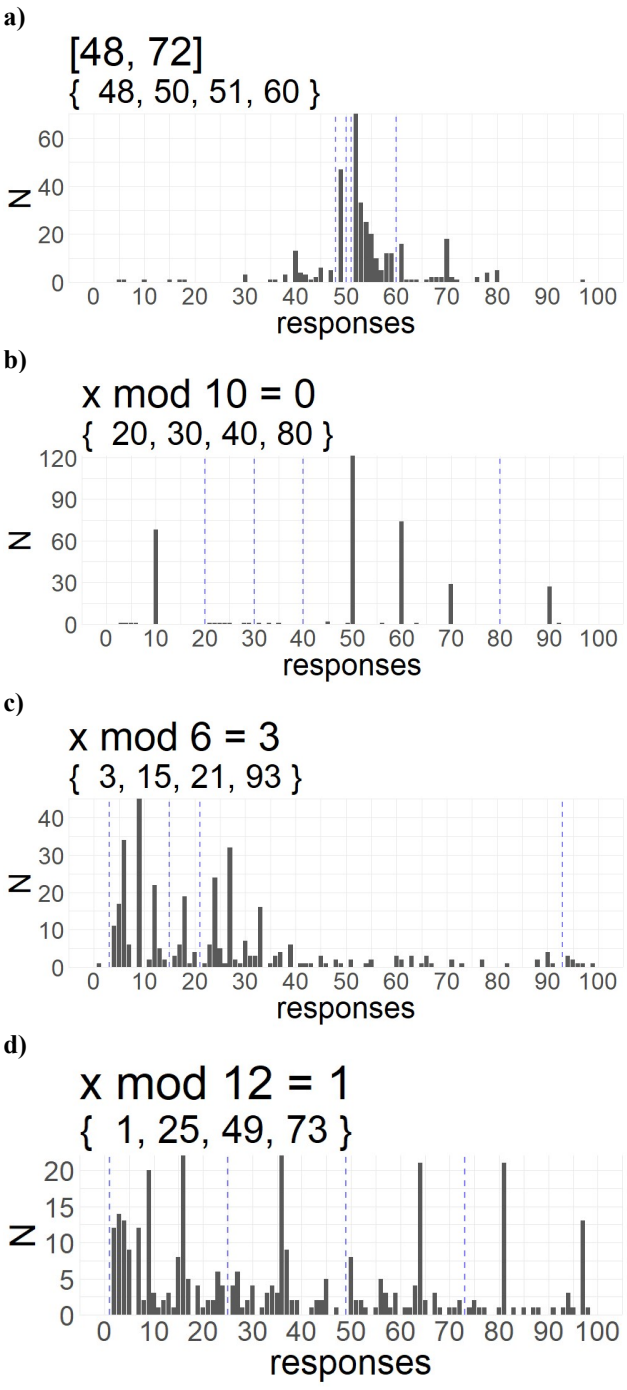

Figure 1: Sample response distributions to the number game generalization task shown in gray. Shown examples are indicated by blue dashed lines. From top to bottom the distributions are representative of the classes: a) interval [48,72]; b) rule multiples [x mod 10=0]; c) rule non-intuitive [x mod 6 = 3] d) rule non-intuitive [x mod 12 = 1].

in generalization (as is observed for gist representations of visual stimuli, for example Whitney, & Yamanashi Leib, 2018).

The novel qualitative results observed in this data likely resulted from the changes in the number game task design relative to Tenenbaum (2000) (e.g., eliciting 2 responses; using less likely and harder to define rules). Overall, however, these empirical distributions are consistent with past work. Thus, we next asked whether these generalization responses predicted memory performance.

Experiment 2: Memory

In Experiment 2, we use the generalization data from Experiment 1 to evaluate whether numbers that were generated more often in response to a set are more likely to be falsely remembered. We rely on our empirical generalization data, rather than a model, to avoid dependence on a particular set of priors and hypotheses (like those formalized by Tenenbaum, 2000). This allows us to ask a more general question about whether propensity to generalize from a set predicts memory errors from that set to that number.

Methods

Participants. A total of 100 participants completed this task. Subjects were included in the final data set if their number of correct responses over the experiment was significantly greater than chance ($p < 0.05$ in a 1 tailed binomial test), which resulted in the exclusion of 9 subjects, leaving a total of 91 participants in the final sample.

Stimuli and Procedure. The same stimuli sets described in Experiment 1 (60 sets of four exemplars) were used as stimuli for memory displays. To create memory errors, we showed the stimuli only briefly. In particular, on each trial, 4 numbers arranged in a circle (randomly placed in one of 8 possible locations) were briefly shown (150ms) and participants were asked to remember the numbers over a delay (2000ms). After the delay, memory was probed in in a 3-alternative forced choice (3-AFC) format, where one choice was the correct item (chosen at random from the 4 encoded items) and the other two choices were sampled from the empirical generalization distributions collected in Experiment 1. The *likely generalization* foil was chosen to be the most frequently generated number for that set of exemplars, and the *unlikely generalization* foil was chosen to be one of the least frequently generated numbers, i.e. at least one subject reported the number in Experiment 1 (in the case of a tie in frequency, the foil was selected at random from the set of numbers with lowest frequency). Choosing an item that was generated at least once provides a conservative test of whether generalization frequency in Exp. 1 predicts memory errors in Exp. 2.

Results and Discussion

Figure 2 shows the proportion of correct responses and the proportion of choices for foils that were *likely generalizations* vs. *unlikely generalizations* in the 3-AFC task. The proportion of correct responses in the memory task was 0.638 (95% CI [0.623, 0.654]), as compared to a chance level of 0.33.

The effect of people's propensity to generalize on memory errors was measured by looking at the proportion of incorrect responses that were to *likely*, as opposed to *unlikely*, foils. A paired sample t-test found the proportion of *likely* responses to be significantly greater than that of *unlikely* responses ($t(90) = 9.058$, $p < 0.001$), meaning that memory errors were more likely when the foil was one of the numbers generated more frequently by participants in Experiment 1.

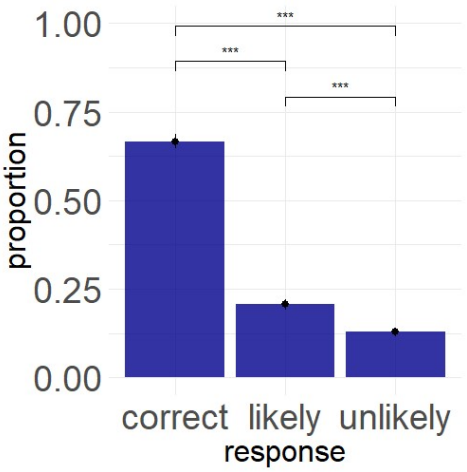

Figure 2: Proportion of responses plotted with 95% confidence intervals. Memory was accurate, and participants chose foils that were *likely generalizations* more than foils that were *unlikely generalizations*.

To test whether this was specific to the shown set, or just because some numbers are frequently produced in both Exp. 1 and 2 regardless of set, we next asked whether this result held when considering only situations where the 'unlikely' foil was generated *more often* in Exp. 1 in sets other than the target set. We found it still held ($t(90) = 4.933$, $p<0.001$).

Was this result only caused by interval sets, where generalization tends to be consistent with simple exponential functions of similarity (e.g., Shepard, 1987)? No: A one-way ANOVA found no significant difference in choosing 'likely' (vs. 'unlikely') foils across the 6 classes of sets, while the intercept was significantly different from 0.5 ($t(53) = 4.668$, $p<0.001$) indicating that for all classes the *likely* foil was chosen more often than the *unlikely* foil. Figure 3 shows the data for the 6 classes of stimuli. This result shows that there was little difference in the effect across different types of sets, so numbers that are likely generalization targets are more likely to be falsely remembered, even when that generalization is based on high level conceptual rules, rather than metric distance to exemplars, and even when the basis of such generalization is not straightforwardly modeled (as in the 'unintuitive' sets).

This provides support for the broad framework of using generalization distributions instead of pairwise similarities in models of memory and suggests formal frameworks for understanding generalization could help make sense of gist-based memory errors.

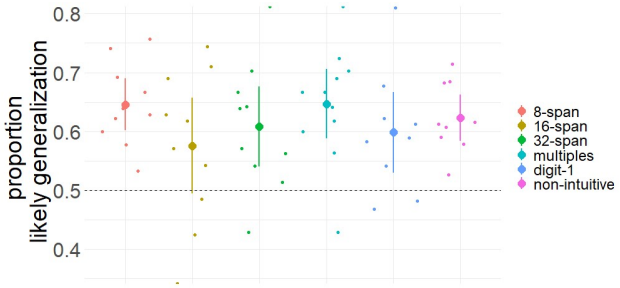

Figure 3: The proportion of incorrect responses that were *likely generalizations* from the examples in Experiment 1. Chance levels correspond to 0.5. The mean effect is plotted with 95% confidence for each class of stimuli.

General Discussion

We proposed a framework for broadening the scope of theories of memory that are based on "similarity". In particular, while previous work has applied formal models of activation and similarity to memory (e.g., Nosofsky, 1992; Schurgin et al. 2020) this approach is significantly limited in its generalizability by its dependence on simple metric similarity structures. Here we propose that such models can be used more broadly to study memory by combining them with a framework for generalization. Our framework suggests a Bayesian model of generalization as the basis for memory activation, which can then be fed into this memory and decision model (e.g., TCC; Schurgin et al. 2020) to predict memory errors in both exemplar-based and rule-based scenarios. We demonstrated the feasibility of this account by empirically connecting the 'number game' generalization task to a task designed to elicit memory errors.

To evaluate our framework, we must consider what the data tell us about the two primary assumptions. The first assumption is that generalization based on concepts or rules will predict memory errors in the same way that similarity in metric spaces predicts memory errors. The data presented provides preliminary support for the idea that generalization in the number game model can predict memory errors. We found that memory errors occurred more often for numbers reported most frequently in the generalization task. This was true across all classes of stimuli, including rule-based situations and non-intuitive rules, suggesting that the distributions of responses generated experimentally in a generalization task predict memory errors even in situations that go well beyond standard conceptions based solely on exponential fall-off in a metric space.

The second assumption of our framework is that observations are used jointly to determine generalization, and thus memory. This assumption has been previously shown in at least some cases of generalization (Tenenbaum & Griffiths, 2001) but is less straightforward to assess with the empirical memory data, and future work will need to fully formalize the proposed framework to better evaluate this assumption. However, some interesting qualitative patterns were observed in Experiment 1. In particular, we see a sort of hybrid generalization behavior in the intuitive rule-based stimuli sets. For example, the set of "multiples of 10" had its most common response near the exemplars; the number of responses that belong to the rule drop off towards the higher end of the number range. Our sampling process for generating exemplar sets was proportional to an exponential distribution, so this pattern seems to approximate that process. One potential issue with creating stimuli in this manner is that the subjects may infer the data generating process which could distort measurement. Further experimentation with different sampling processes is needed to determine the limitations this could impose on our framework.

Another interesting qualitative finding comes from the patterns observed in the response distributions for "non-intuitive" rule-based exemplar sets. These sets exhibited properties similar to visual ensemble processing, where "outliers" are discounted in the gist representation (see Whitney, & Yamanashi Leib, 2018), broadly consistent with weighted cue combination (Landy et. al. 1995). This pattern provides some support for the idea that multiple examples jointly determine generalization.

The main limitation of this study is that empirical data cannot provide unequivocal support for the theoretical framework, only preliminary evidence. Future work should focus on formalizing a model of memory errors through the framework proposed — with memory responses explicitly arising from noise-perturbed activations derived from a Bayesian model of generalization. Our experimental paradigms demonstrate the plausibility of this model and the qualitative patterns observed in the data can provide future directions of research. The next step for this project is to develop and implement a computational model that can be applied to data collected in these paradigms and design experiments within them that test specific predictions of the model. The broad goal of our framework is to create a unifying theory of memory that can capture behavior in both low- and high-level domains of cognition.

References

Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of mathematical psychology*, *39*(2), 216-233.

Bartlett, F C. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press.

Gallo, D. . (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*(7), 833-848

Goldstone R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. Journal of memory and language, 37(4), 555-583.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. Vision research, 35(3), 389-412.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*(1), 39.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual review of Psychology*, *43*(1), 25-53.

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.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature human behaviour*, *4*(11), 1156-1172.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. Science, 210(4468), 390-398.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317-1323.

Tenenbaum, J. B. (2000). Rules and similarity in concept learning. *Advances in neural information processing systems*, *12*, 59-65.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and brain sciences*, *24*(4), 629.

Tversky, A. (1977). Features of similarity. *Psychological review*, *84*(4), 327.

Van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780-8785.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual review of psychology*, *69*, 105-129.