Comparisons promote a tradeoff between generalization and individuation

Nicholas Ichien
ichien@ucla.edu

Angela Kan
kangela@ucla.edu

Department of Psychology
University of California, Los Angeles
Los Angeles, CA 90095 USA

Abstract

Qualitative comparison provides learning opportunities that support a range of cognitive abilities. The present research clarifies how the contrasting goals of qualitative comparison, processing similarities versus differences, contribute to the overall impact that comparison has on learning. We argue that attending to similarities produces memory traces that are dissociable from those left by attending to differences, and that these memory traces reflect a tradeoff between generalization and individuation. Across two preregistered experiments, participants either attended to similarities or differences between stimuli. We show that whereas attending to similarities yielded better performance on a category recognition task (Experiment 1), attending to differences yielded better performance on an old/new recognition task (Experiment 2). Comparison thus serves as a learning opportunity that enables agents to manage the generalization-individuation tradeoff by shifting their attention toward similarities or differences.

Keywords: concepts and categories, learning, memory, representation

Introduction

The cognitive act of qualitative comparison serves at least two immediate goals: To process similarities between whatever is compared, and to process differences between them. Whereas past work has clarified how people carry out this processing (Golonka & Estes, 2009; Ichien et al., 2024; Markman & Gentner, 1993; Medin et al., 1990; Sagi et al., 2012; Tversky, 1977; Tversky & Gati, 1978), the present work examines the *consequences* of this processing. In general, comparison promotes human learning that is applicable in a range of downstream cognitive abilities, including classification (Jacoby et al., 2024; Kurtz et al., 2013; Spalding & Ross, 1994), language use (Gentner & Namy, 2006; Goldwater, 2017), and problem solving (Gick & Holyoak, 1983). And these learning benefits hold in more naturalistic educational settings (Alfieri et al., 2013). The aim of the present research is to clarify how the contrasting goals of processing similarities and differences respectively contribute to the overall impact that comparison has on learning.

As a starting point for the present analysis of similarity and difference, we rely on a distinction between 'within-category' and 'between-category' comparisons in supervised category learning (e.g., Corral et al., 2018; Goldstone, 1996; Higgins & Ross, 2011). In a typical category learning experiment, participants are presented with a series of stimuli, each paired with a category label, and are tasked with learning how to correctly classify stimuli according to their labels. Within-category comparisons are between stimuli with the same category label, and between-category comparisons are between stimuli with different category labels. These kinds of comparisons differ not only in the input stimuli over which they operate but also in their immediate goals: Within-category comparisons are oriented toward finding similarities between stimuli with a common label, whereas between-category comparisons are oriented toward finding differences between stimuli with distinct labels. To the extent that within- and between-category comparisons make distinct contributions to learning, so too might similarity and difference processing.

The clearest dissociation between within- and between-category comparisons involves effects of stimulus sequencing in category learning (for other dissociations between within- and between-category comparisons based on simultaneous presentations of stimulus pairs, see Corral et al., 2018; Higgins & Ross, 2011). In 'blocked' sequencing, stimuli are ordered according to their category label (e.g., A A A B B B C C C), whereas in 'interleaved' sequencing, stimuli are shuffled according to their label (e.g., A C B A B C B A C). Sometimes, blocked sequencing promotes more accurate classification performance than does interleaved (Sorensen & Woltz, 2016), but in other cases the reverse is true (Kornell & Bjork, 2008). One explanation for these seemingly conflicting results is rooted in the observation that on one hand, blocked sequencing affords more within-category comparisons and thus more opportunities to selectively attend to diagnostic features of categories between consecutively-presented stimuli; whereas on the other hand, interleaved sequencing affords more between-category comparisons and thus more opportunities to selectively attend to features that discriminate between categories (Carvalho & Goldstone, 2014; Goldstone, 1996; for a different explanation, see Vlach, 2014; Vlach et al., 2008).

In line with this proposal, human learners exhibit a blocking advantage in acquiring categories with low within- and between-category similarity, where opportunities to attend to diagnostic features are more valuable than are opportunities to attend to discriminating features; but they exhibit an interleaving advantage in acquiring categories with high within- and between-category similarity, where the reverse is true (Brunmair & Richter, 2019; Carvalho & Goldstone, 2014).

This attention-driven complementarity between within- and between-category comparisons suggests a corresponding

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

complementarity between comparisons oriented toward similarities and those oriented toward differences. Outside of the context of explicit category learning, comparisons aimed at similarities might simply prompt attention to features that are shared among stimuli, whereas comparisons aimed at difference prompt attention to features that discriminate stimuli. We thus hypothesize that similarity and difference encode memory traces that respectively promote generalization and individuation, similar to other previously-noted dichotomies such as relational versus item-specific encoding (Einstein & Hunt, 1980; Hunt & Einstein, 1981), gist versus verbatim memory traces (Brainerd & Reyna, 2002), and categorization versus identification (Bruner, 1956; Nosofsky, 1986), as well as a dissociation between processing of quantifier terms “every” and “each” in psycholinguistics (Knowlton et al., 2023).

The cognitive operations that an agent performs on a stimulus to encode that stimulus are known to constrain the retrieval conditions that effectively facilitate retrieving the corresponding memory trace (Craik & Tulving, 1975; Morris et al., 1977; but for a critique of the notion that this constraint is rooted in similarity between encoding and retrieval conditions, see Nairne, 2002). We adopted this *transfer-appropriate processing* framework to test the hypothesized complementarity between similarity and difference. In two preregistered experiments, we manipulated the goal of comparison to vary encoding conditions for pairs of stimuli (i.e., to attend to similarities between stimuli or to attend to differences; see Figure 1). We then compared recognition memory under retrieval conditions that respectively prioritize generalization (‘category recognition’ in Experiment 1), for which we predict superior performance following similarity than difference comparisons; and individuation (‘old/new recognition’ in Experiment 2), for which we predict the reverse.

Experiment 1

In a first experiment, we assessed comparisons oriented toward similarity and those oriented toward difference as encoding conditions for memory traces that promote generalization. We presented participants with pairs of paintings, and either asked them to identify similarities between the paintings or to identify differences between them. Following past work showing that people are better at identifying differences between similar stimuli, relative to dissimilar stimuli (Gentner & Markman, 1994; Sagi et al., 2012), we restricted stimulus pairs to paintings painted by the same artist. Then, after a distractor task, we asked them to judge whether novel paintings were painted by the same artist as any of the artists who painted those they had seen earlier (for a similar task, see Experiment 2 of Kornell & Bjork, 2008). We hypothesized that by directing participants’ attention to common features among paintings, comparisons oriented toward similarities would promote better generalization performance than would comparisons oriented toward differences, which we hypothesized would direct participants’ attention to distinct features among paintings. In order to ensure that the encoding conditions of task stimuli more purely consisted of stimulus comparison, rather than being contaminated by participants’ intention to remember (Popov & Dames, 2023), we adopted an incidental memory paradigm in which inclusion of a memory task was only revealed to participants after their distractor task. The hypotheses, methods, and analysis plan of category recognition data were preregistered (https://osf.io/9aeh4).

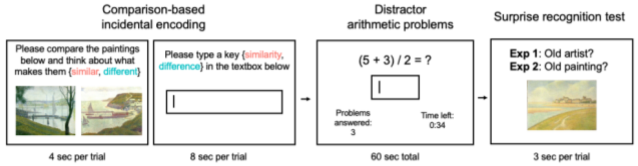

Figure 1: Experimental tasks. Both Experiments 1 and 2 consisted of a comparison-based incidental encoding task (leftmost panels), coupled with distractor arithmetic problems (middle panel), but differed in their recognition memory task (rightmost panels). Experiment 1 used a category recognition memory task, and Experiment 2 used an old/new recognition memory task.

Method

Participants We recruited a total of 352 participants from Prolific Academic based on a power analysis, which would enable us to detect a between-subject effect with a minimum effect size, as measured by Cohen’s d, $d = .3$, with statistical power, $\beta = .80$. We limited our data collection to participants who were fluent in English and had an approval rate above 98% on Prolific. We estimated that the task would take 10 minutes total, and we paid participants $2 for their completion (which corresponds to a rate of $12 per hour). Research procedures were approved by the Institutional Review Board at the University of Pennsylvania (the first author’s former institution).

Materials and procedure Participants completed three tasks in a fixed order; Figure 1 shows example trials from each of these tasks. The experimental session started with a task intended to prompt incidental encoding of task stimuli via comparison (Figure 1, left panels); continued with a series of basic arithmetic problems, which served as a distractor task (Figure 1, middle panel); and ended with a surprise task testing recognition memory for stimulus categories, which requires selective generalization from previously encoded stimuli to novel stimuli (Figure 1, upper-right panel). All tasks were designed using Psychopy software (Peirce et al., 2019) and were administered online via Pavlovia (https://pavlovia.org/).

Comparison-based incidental encoding Participants’ experimental session started with a task intended to prompt incidental encoding of task stimuli. On each trial of this task, participants were first shown a pair of paintings, each painted by the same artist, for 4 seconds (Figure 1, leftmost panel). Prior to the experiment, participants were randomly assigned to one of two conditions. In the similarity condition, participants were instructed to compare each pair of paintings and think about what makes them similar; in the difference condition, participants were instructed to compare each pair of paintings and think about what makes them different. Participants were then given 8 seconds to type in a textbox a key similarity or difference between the paintings they just saw, depending on their assigned condition (Figure 1, second panel from the left). After those 8 seconds, the textbox disappeared, and a new pair of paintings appeared.

Overall, participants completed 12 trials of this task; since participants were shown two paintings on each trial, participants were shown 24 paintings total. Across their 12 trials, 2 trials presented unique paintings painted by the same artist; hence a given participant saw 4 unique paintings painted by the same artist (i.e., 4 paintings by each of 6 different artists).

Painting stimuli were drawn from a larger stimulus set introduced by Kornell and Bjork (2008) and made publicly available by Nate Kornell (https://sites.williams.edu/nk2/stimuli/). These stimuli consist of 10 unique paintings painted by each of 12 different artists (120 paintings total): Georges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pessani, Ron Schlorff, Georges Seurat, Ciprian Stratulat, George Wexler, and YieMei. Which 6 of these artists' paintings, and which 4 of their paintings were shown to a given participant during this task, was counterbalanced across participants. Painting pairs were shown in a random order.

Distractor arithmetic problems Next, participants were asked to solve as many arithmetic problems as they could in 1 minute. Following Popov and Dames (2023), these problems consisted of three one- or two-digit integers, X, Y, and Z, and required either adding Y to X or subtracting Y from X in parentheses, and then multiplying or dividing that value by Z (e.g., $(X + Y) / Z = ?$; Figure 1 middle panel). These problems were drawn from a list of 30 total problems and were presented in a random order. Participants were asked to type their response in a textbox and then press 'Enter' to submit their response, after which the next problem would appear. The correct response to each of these problems was a positive or negative integer.

Even though participants were not told about the memory task that they would complete next, these arithmetic problems served as a buffer between encoding and memory tasks to reduce the likelihood that participants might maintain representations of encoded stimuli in working memory during the memory task.

Surprise category recognition memory Participants' experimental tasks concluded with a surprise memory task. Participants were informed of this task only after they completed their minute-long distractor task. They were told that they would be presented with a series of 24 completely new paintings, none of which they had been shown during the prior encoding task. On each trial, participants were presented with a single panting, and they were instructed to indicate whether that painting was painted by one of the same artists that painted any of the paintings they had seen during the encoding task. Each painting was shown for 3 seconds, during which participants could use the 'p' key to indicate 'Yes' (i.e., the artist who painted this painting also produced one of the paintings shown during the encoding task) or use the 'q' key to indicate 'No' (Figure 1, upper-right panel). They were not asked to identify the artist.

Of the 24 novel paintings displayed during this task, 12 were painted by one of the six artists whose work was shown during the previous comparison task (i.e., 'familiar'), so the correct response for these trials was 'Yes'. The 12 remaining paintings were painted by six other artists whose work was not shown during the comparison task (i.e., 'unfamiliar'), so the correct response for these trials was 'No'. Painting stimuli were presented in a random order across participants.

Table 1: Participant-generated similarities (above) and differences (below) of depicted pair of paintings. Each bullet point includes a description from a different participant.

similarities

- "both paintings have boats and water"
- "Green Water"
- "pointillism"
- "Still water with land beside"
- "The vibes, and the greeny coloring very grainy and all are ships" [*sic*]

differences

- "The first one is colder and darker"
- "lonely harbour vs fishing village looks"
- "dark colors other light"
- "one painting is sad sea the other is nice town"
- "one looks slightly black and wite" [*sic*]

Results and Discussion

Before reporting analyses of category recognition memory performance, we examine painting descriptions produced during comparison-based encoding in order to validate our manipulation of comparison type.

Painting descriptions (from Comparison-based incidental encoding) Recall that during their encoding task, participants were shown pairs of paintings and were instructed to either type a similarity or a difference between those two paintings, depending on their experimental condition. We first sought to assess whether participants followed task instructions such that participants asked to type out similarities did, in fact, ype out similarities, and that participants asked to type out differences did type out differences. Table 1 shows five participant-generated similarities and five participant-generated differences of the depicted pair of paintings, each produced by a different participant. All participant-generated descriptions are available at the OSF page for this project (https://osf.io/bs8tx/?view_only=63e277c727404df1ab4e9bd21ed2138e)

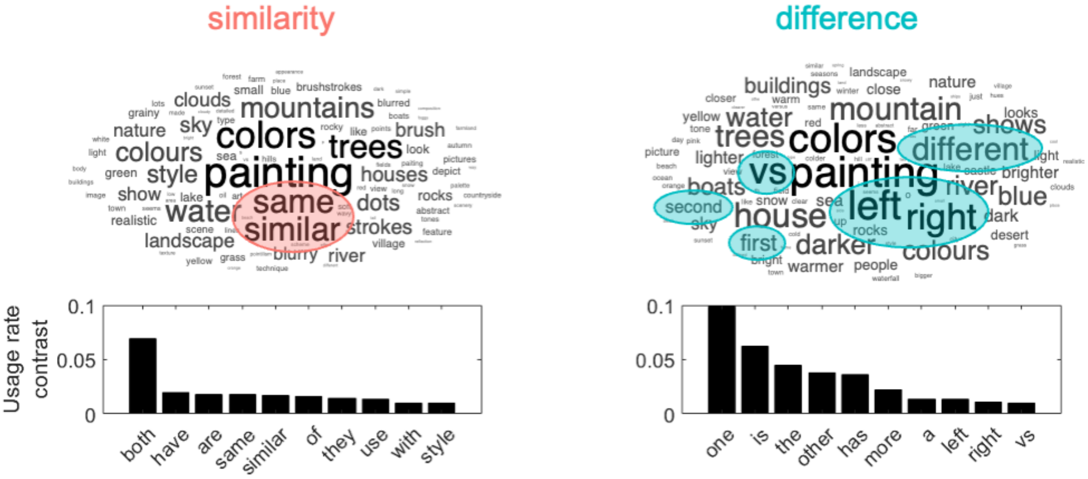

Figure 2: Most frequent words appearing in painting descriptions, broken down by condition. Circled words within each word cloud were selected as highly reflective of their experimental condition. Bar graphs below show words with the greatest difference in frequency across conditions. Left graph show words that were more frequent in similarity descriptions than in difference descriptions, and vice versa for right graph.

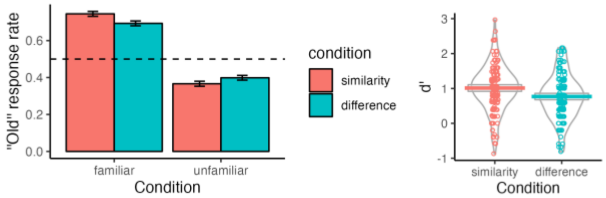

Figure 3: Experiment 1 Category recognition memory performance, broken down by condition. Left panel shows rate of indicating that a given painting was painted by an artist whose paintings were shown during the encoding task (dashed line indicates chance response rate), and right panel shows *d'*.

In order to provide an overall impression of these descriptions, we focus discussion on the most frequently used words in participant-generated descriptions. No analyses of participant-generated descriptions were preregistered. Figure 2 depicts the most frequent words used in similarities (left) and differences (right). Word clouds show that both descriptions tend to use many overlapping words, with notable exceptions: "same" and "similar", used in similarities, and "different", "vs", "left", "right", "first", and "second", used in differences. We identified what words were used more often in similarities than in differences (and vice versa) by finding words whose rates of use had the highest contrast across conditions. We computed these usage rates for each unique word mentioned in participant-generated descriptions and did so separately for each condition. We scaled the frequency that each unique word was used by the total number of words used across descriptions and then ranked words according to those whose usage rates had the highest contrast across conditions (i.e., "usage rate contrast"; see Figure 2 bar graphs). Among these, words appearing more often in similarities than in differences included "both", used to refer to the two paintings together, plural-conjugated verbs (e.g., "have", "are", "use"), as well as the words "same", and "similar". In contrast, words appearing more often in differences than in similarities included "one", "other", "left", "right", used to single out one of the paintings, and singular-conjugated verbs (e.g., "has", "is"), as well as "vs", which expresses a contrast. This qualitative analysis of participant-generated descriptions validates that participants engaged in the encoding task as instructed.

Category recognition memory performance. Overall, participants did well on this task across conditions (*similarity*: M_{acc} = .68, SD_{acc} = .13; *difference*: M_{acc} = .64, SD_{acc} = .12). The left panel of Figure 3 shows hit rates (*similarity*: M_{hit} = .71, SD_{hit} = .17; *difference*: M_{hit} = .67, SD_{hit} = .16), and the middle panel shows false alarm rates (*similarity*: $M_{false\text{-}alarm}$ = .36, $SD_{false\text{-}alarm}$ = .16; *difference*: $M_{false\text{-}alarm}$ = .40, $SD_{false\text{-}alarm}$ = .16) broken down by condition. We examined whether identifying similarities yielded better category recognition memory than did identifying differences. To do so, we computed *d'* for each participant to measure individual-level memory sensitivity, and then we compared mean *d'* values for each condition, which is shown on the right panel of Figure 3. To address extreme values in computing *d'* (i.e., hit rates or false alarm rates of 1 or 0), we adopted the log-linear method, which adjusts all hit rates and false alarm rates according to the following rule: $rate_{adjusted} = rate_{raw} + \frac{1}{2(n+1)}$, where n = number of old / signal-present trials (for hit rate) or new / signal-absent trials (n = 12 for both). Monte Carlo simulations have shown that the log-linear method produces less biased estimates of *d'* than the popular approach of respectively adding $\frac{1}{2n}$ to or subtracting it from hit rates or false alarm rates of 0 or 1 (i.e., the '1/2*n* rule') (Hautus, 1995).

Independent-samples t-tests showed that participants attending to similarities had greater memory sensitivity ($M_{d'}$ = 1.01, $SD_{d'}$ = .64) than did participants attending to differences ($M_{d'} = .77$, $SD_{d'} = .63$), $t(348.72) = 3.60$, $p < .001$, $d = .39$, and conditions did not differ in their criterion / bias (*similarity*: $M_{bias} = -.12$, $SD_{bias} = .40$; *difference*: $M_{bias} = .11$, $SD_{bias} = .37$), $t(346.27) = -.42$, $p = .67$. This result held even after excluding participants with negative *d*' values, $t(312.83) = 3.61$, $p > .001$, $d = .41$, as well as when we compared *A'* values, a nonparametric measure of memory sensitivity (Sloutsky & Fisher, 2004; Wickens, 2001), $t(348.76) = 2.85$, $p = .005$, $d = .30$ (*similarity*: $M_{A'} = .74$, $SD_{A'} = .16$; *difference*: $M_{d'} = .69$, $SD_{A'} = .16$), and overall accuracy (see above for means and standard deviations), $t(344.59) = 2.83$, $p = .005$, $d = .30$.

In order to further specify the source of this performance gap across similarity and difference participants, we fit a logistic mixed-effects model of trial-level accuracy on the category recognition memory task using the *lme4* package in R (Bates et al., 2015). This model included a *condition* (*similarity* versus *difference*) by *stimulus type* (*familiar* versus *unfamiliar*) interaction term, and *1+stimulus type | participant* and *1 | stimulus* random-effect terms. In an effort to bolster the generalizability of our analysis (Barr et al., 2013), we used the *buildmer* R package to determine the maximal random-effect structure of our model that would converge.

A likelihood-ratio test compared this full model to a reduced model that lacked the interaction term, but that was an otherwise equivalent model. Though the interaction term did not contribute to overall model fit, Δ AIC = 1, $\chi^2 = .68$, $p = .41$, we nevertheless examined marginal means estimated from this full model in order to test simple-effects of condition within each stimulus type. Similarity participants recognized 'familiar' paintings (i.e., paintings created by artists whose paintings were shown during the encoding task) at a higher rate (i.e., had higher hit rates, see Figure 1 left panel) than difference participants, *estimated mean difference* = .22, $SE = .08$, $z = 2.61$, $p = .009$. For 'unfamiliar' paintings (i.e., paintings created by artists whose paintings were not shown during the encoding task), there was no difference between the two conditions (i.e., in false alarm rates, see Figure 1 middle panel), *estimated mean difference* = .12, $SE = .08$, $z = 1.52$, $p = .13$. These results suggest that attending to similarities promotes better generalization than does attending to differences.

Experiment 2

Experiment 1 showed that attending to similarities between stimuli in within-category comparisons prompts more accurate generalization to novel instances of those categories than does attending to differences between those same stimuli. This result supports the hypothesis that selective attention to similarities between stimuli aids generalization across those stimuli. However, features that benefit category recognition likely differ from features important for recognizing individuals (De Brigard et al., 2017; Meagher & Nosofsky, 2023; Nosofsky, 1991). Correspondingly, similarity-comparisons might be expected to *harm* task performance on old/new recognition memory, where more granular representations of individuals are emphasized (for similar approaches, see Sloutsky & Fisher, 2004; Vendetti et al., 2014). In Experiment 2 we adopted the same procedure as Experiment 1, but replaced the category recognition task with an old/new recognition memory task, in which participants were asked to judge whether they had seen specific paintings during the previous encoding task. Experiment hypotheses, methods, and analysis plan for old/new recognition data were preregistered (https://osf.io/9aeh4).

Method

Participants We recruited a total of 352 new participants from Prolific Academic. This sample size was determined by the same power analysis used in Experiment 1. We limited our data collection to participants who were fluent in English and had an approval rate above 98%. Participants were only allowed to participate once and earned a payment equal to $12 an hour. Research procedures were approved by the Institutional Review Board at the University of Pennsylvania.

Procedure Experimental procedure was identical to Experiment 1, except the surprise category recognition task was replaced with a surprise old/new recognition memory task.

Surprise old/new recognition memory As in Experiment 1, participants in Experiment 2 concluded their experimental sessions with a surprise memory task of which they were not informed until after completing their distractor task. Participants were told that they would be presented with a series of 36 paintings, some of which they were shown during the comparison task that they had completed earlier. As in Experiment 1, paintings were presented one at a time, and participants were instructed to indicate which paintings they had seen during the comparison task. Unlike the category recognition task used in Experiment 1, artists were not mentioned in the instructions in this old/new recognition memory task. Each painting was displayed for 3 seconds, during which participants could use the 'p' key to indicate 'Yes', this painting was shown during the comparison task, or use the 'q' key to indicate 'No'. Stimulus order was randomized across participants.

Of the 36 paintings displayed during this task, 12 were paintings shown during the encoding task (i.e., 'old'), so the correct response for these trials was 'Yes'. The remaining 24 paintings were novel, so the correct response for these trials was 'No'. Of these paintings, 12 were painted by one of the six artists whose work was shown during the previous comparison task (i.e., 'familiar'), and the other 12 were painted by six other artists whose work was not shown during the comparison task (i.e., 'unfamiliar').

Results

Painting descriptions (from Comparison-based incidental encoding) The results from Experiment 2 were practically identical to those in Experiment 1, again confirming that participants followed task instructions during their encoding task.

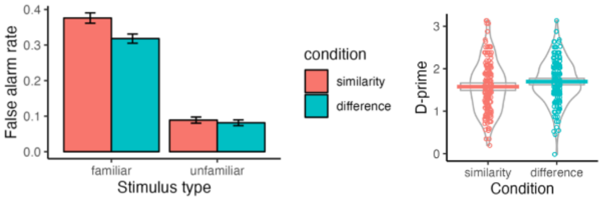

Figure 4: Experiment 2 old/new recognition memory performance, broken down by condition. Left panel shows false alarm rates on ‘familiar’ and ‘unfamiliar’ paintings, and right panel shows *d’*.

Old/new recognition memory performance As in Experiment 1’s category recognition task, participants did well on this task across conditions (*similarity*: M_{acc} = .78, SD_{acc} = .09; *difference*: M_{acc} = .80, SD_{acc} = .08). Participants had high hit rates (*similarity*: M_{hit} = .76, SD_{hit} = .13; *difference*: M_{hit} = .77, SD_{hit} = .13) and moderately low false alarm rates (*similarity*: $M_{false\text{-}alarm}$ = .24, $SD_{false\text{-}alarm}$ = .12; *difference*: $M_{false\text{-}alarm}$ = .21, $SD_{false\text{-}alarm}$ = .10), as shown in the left panel of Figure 4.

We examined whether attending to differences yielded better recognition memory than did attending to similarities. To do so, we compared memory sensitivity as computed by *d’* values for each participant, shown in the right panel of Figure 4. As in Experiment 1, we used the log-linear method to address extreme values (i.e., hit rates or false alarm rates of 1 or 0) (Hautus, 1995). Participants attending to differences ($M_{d'}$ = 1.70, $SD_{d'}$ = .50) had greater memory sensitivity than did participants attending to similarities ($M_{d'}$ = 1.58, $SD_{d'}$ = .58), $t(342.13) = 2.07$, $p = .04$, $d = .22$, and conditions did not differ in their criterion / bias (*similarity*: M_{bias} = .00, SD_{bias} = .35; *difference*: M_{bias} = .03, SD_{bias} = .37), $t(349.06) = .77$, $p = .44$. This result held after excluding participants with negative *d*’ values, $t(338.41) = 2.26$, $p = .02$, $d = .24$, as well as for *A’* values (*difference*: $M_{A'}$ = .86, $SD_{A'}$ = .06; *similarity*: $M_{d'}$ = .84, $SD_{A'}$ = .07), $t(342.61) = 2.48$, $p = .01$, $d = .27$, and overall accuracy (see above for means and standard deviations), $t(341.49) = 2.89$, $p = .004$, $d = .31$.

In order to further specify the source of this performance gap across similarity and different participants, we fit a logistic mixed-effects model of trial-level accuracy on the old / new recognition memory task using the *lme4* package in R (Bates et al., 2015). This model included a *condition* (*similarity* versus *difference*) by *stimulus type* (*old* versus *familiar* versus *unfamiliar*) interaction term, and *1 | participant* and *1 | stimulus* random-effect terms. In order to improve the generalizability of our analysis (Barr et al., 2013), we used the *buildmer* R package to determine the maximal random-effect structure of our model that would converge.

Although the interaction term did not contribute to overall model fit (Δ AIC = 0, $\chi^2 = 4.50$, $p = .11$), we examined marginal means estimated from this full model to test simple effects of condition within each stimulus type. Similarity participants misrecognized ‘familiar’ paintings at a higher rate (i.e., had a higher false alarm rate, see Figure 4 middle panel) than difference participants, *estimated mean difference* = .28, *SE* = .08, $z = 3.60$, $p = .003$. There was no difference across conditions for ‘unfamiliar’ paintings, *estimated mean difference* = .09, *SE* = .12, $z = .79$, $p = .43$, or on ‘old’ paintings, *estimated mean difference* = .07, *SE* = .09, $z = .86$, $p = .39$. Together with the results of Experiment 1, these findings suggest that attending to similarities between exemplars promoted stimulus generalization to the *detriment* of stimulus individuation in Experiment 2.

General Discussion

Across two preregistered experiments, we showed that the goals of comparison systematically constrain the retrieval conditions that are most compatible with the memory traces left by that comparison. Relative to differences, attending to similarities was more compatible with retrieval conditions emphasizing stimulus generalization in a category recognition task (Experiment 1). The reverse was found in a recognition memory task (Experiment 2), for which accuracy was greater when participants attended to differences and retrieval conditions emphasized stimulus individuation. These results demonstrate a complementarity between the consequences of similarity and difference processing, which mirrors a number of previously-observed dichotomies observed in human cognition (Brainerd & Reyna, 2002; Bruner, 1956; Hunt & Einstein, 1981, 1981; Knowlton et al., 2023; Nosofsky, 1986).

The present results also align with demonstrations of similarity-induced memory biases in working memory (Fukuda et al., 2022; Saito et al., 2023). Exposure to a perceptually available stimulus that is judged to be similar to a color representation held in working memory biases people’s working memory representations to be more similar to that stimulus than does exposure to a stimulus that is judged to be dissimilar. On the other hand, Fukuda et al. did not find evidence for working memory biases away from the presented stimulus, relative to a baseline condition, when the stimulus was judged to be dissimilar. In some cases, participants even exhibited biases *toward* stimuli judged to be dissimilar. This asymmetry between subjectively similar and dissimilar stimuli raises the possibility that difference-comparisons might promote generalization relative to some baseline condition. Correspondingly, in cases of category-induced representational change, category learning produces within-category representational *compression* such that stimuli falling under the same label come to be seen as more similar to each other (Dubova & Goldstone, 2022; Goldstone et al., 2001; Livingston et al., 1998), but not between-category representational *repulsion* (Goldstone et al., 2001). Future work is needed to further clarify the impact of difference comparisons on human learning and memory.

Overall, the present experiments demonstrate that the contrasting goals of processing similarities and processing differences produce memory traces that reflect a trade-off between generalization and individuation. Comparison thus serves as a learning opportunity that enables learners to navigate this trade-off by strategically shifting their attention to similarities or differences.

Acknowledgments

We thank Keith J. Holyoak, David J. Halpern, Julia J. Conti, Sara D. Jaramillo, members of Oak lab at Carnegie Mellon University, and members of the "Meso" Cognitive Science group at Indiana University for helpful feedback and discussions. The work was supported by National Science Foundation SBE Postdoctoral Research Fellowship 2313985 awarded to Nicholas Ichien.

References

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist*, *48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*(5), 164–169. https://doi.org/10.1111/1467-8721.00192

Bruner, J. (1956). *A Study of Thinking* (1st ed.). Wiley.

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*(11), 1029–1052. https://doi.org/10.1037/bul0000209

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

Corral, D., Kurtz, K. J., & Jones, M. (2018). Learning relational concepts from within- versus between-category comparisons. *Journal of Experimental Psychology. General*, *147*(11), 1571–1596. https://doi.org/10.1037/xge0000517

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

De Brigard, F., Brady, T. F., Ruzic, L., & Schacter, D. L. (2017). Tracking the emergence of memories: A category-learning paradigm to explore schema-driven recognition. *Memory & Cognition*, *45*(1), 105–120. https://doi.org/10.3758/s13421-016-0643-6

Dubova, M., & Goldstone, R. L. (2022). Categorical perception meets El Greco: Categories unequally influence color perception of simultaneously present objects. *Cognition*, *223*, 105025. https://doi.org/10.1016/j.cognition.2022.105025

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 588–598. https://doi.org/10.1037/0278-7393.6.5.588

Fukuda, K., Pereira, A. E., Saito, J. M., Tang, T. Y., Tsubomi, H., & Bae, G.-Y. (2022). Working memory content Is distorted by its use in perceptual comparisons. *Psychological Science*, *33*(5), 816–829. https://doi.org/10.1177/09567976211055375

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *5*(3), 152–158. https://doi.org/10.1111/j.1467-9280.1994.tb00652.x

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Directions in Psychological Science*, *15*(6), 297–301. https://doi.org/10.1111/j.1467-8721.2006.00456.x

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1–38. https://doi.org/10.1016/0010-0285(83)90002-6

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*(5), 608–628. https://doi.org/10.3758/BF03201087

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition*, *78*(1), 27–43. https://doi.org/10.1016/S0010-0277(00)00099-8

Goldwater, M. B. (2017). Grammatical constructions as relational categories. *Topics in Cognitive Science*, *9*(3), 776–799. https://doi.org/10.1111/tops.12272

Golonka, S., & Estes, Z. (2009). Thematic relations affect similarity via commonalities. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(6), 1454–1464. https://doi.org/10.1037/a0017397

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values ofd′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. https://doi.org/10.3758/BF03203619

Higgins, E. J., & Ross, B. H. (2011). Comparisons in category learning: How best to compare for what. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *33*.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 497–514. https://doi.org/10.1016/S0022-5371(81)90138-9

Ichien, N., Lin, N., Holyoak, K. J., & Lu, H. (2024). Cognitive complexity explains processing asymmetry in judgments of similarity versus difference. *Cognitive Psychology*, *151*, 101661. https://doi.org/10.1016/j.cogpsych.2024.101661

Jacoby, V. L., Massey, C. M., & Kellman, P. J. (2024). Adaptively triggered comparisons enhance perceptual category learning: Evidence from face learning. *Scientific Reports*, *14*(1), 19814. https://doi.org/10.1038/s41598-024-70163-6

Knowlton, T., Trueswell, J., & Papafragou, A. (2023). Keeping quantifier meaning in mind: Connecting semantics, cognition, and pragmatics. *Cognitive Psychology*, *144*, 101584. https://doi.org/10.1016/j.cogpsych.2023.101584

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kurtz, K. J., Boukrina, O., & Gentner, D. (2013). Comparison promotes learning and transfer of relational categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1303–1310. https://doi.org/10.1037/a0031847

Livingston, K. R., Andrews, J. K., & Harnad, S. (1998). Categorical perception effects induced by category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *24*(3), 732–753. https://doi.org/10.1037//0278-7393.24.3.732

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, *32*(4), 517–535. https://doi.org/10.1006/jmla.1993.1027

Meagher, B. J., & Nosofsky, R. M. (2023). Testing formal cognitive models of classification and old-new recognition in a real-world high-dimensional category domain. *Cognitive Psychology*, *145*, 101596. https://doi.org/10.1016/j.cogpsych.2023.101596

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science*, *1*(1), 64–69. https://doi.org/10.1111/j.1467-9280.1990.tb00069.x

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory*, *10*(5–6), 389–395. https://doi.org/10.1080/09658210244000216

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology. General*, *115*(1), 39–61. https://doi.org/10.1037//0096-3445.115.1.39

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(1), 3–27. Scopus. https://doi.org/10.1037/0096-1523.17.1.3

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Popov, V., & Dames, H. (2023). Intent matters: Resolving the intentional versus incidental learning paradox in episodic long-term memory. *Journal of Experimental Psychology: General*, *152*(1), 268–300. https://doi.org/10.1037/xge0001272

Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science*, *36*(6), 1019–1050. https://doi.org/10.1111/j.1551-6709.2012.01250.x

Saito, J. M., Duncan, K., & Fukuda, K. (2023). Comparing visual memories to similar visual inputs risks lasting memory distortion. *Journal of Experimental Psychology: General*, *152*(8), 2318–2330. https://doi.org/10.1037/xge0001400

Sloutsky, V. M., & Fisher, A. V. (2004). When development and learning decrease memory: Evidence against category-based induction in children. *Psychological Science*, *15*(8), 553–558. https://doi.org/10.1111/j.0956-7976.2004.00718.x

Sorensen, L. J., & Woltz, D. J. (2016). Blocking as a friend of induction in verbal category learning. *Memory & Cognition*, *44*(7), 1000–1013. https://doi.org/10.3758/s13421-016-0615-x

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(6), 1251–1263. https://doi.org/10.1037//0278-7393.20.6.1251

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352. https://doi.org/10.1037/0033-295X.84.4.327

Tversky, A., & Gati, I. (1978). Studies of similarity. In *Cognition and categorization* (pp. 79–98). Lawrence Erlbaum.

Vendetti, M. S., Wu, A., Rowshanshad, E., Knowlton, B. J., & Holyoak, K. J. (2014). When reasoning modifies memory: Schematic assimilation triggered by analogical mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1172–1180. https://doi.org/10.1037/a0036350

Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163–168. https://doi.org/10.1111/cdep.12079

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167. https://doi.org/10.1016/j.cognition.2008.07.013

Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford University Press.