

How similar are recognition memory and inductive reasoning?

Brett K. Hayes · Evan Heit

Published online: 31 January 2013
© Psychonomic Society, Inc. 2013

Abstract Conventionally, memory and reasoning are seen as different types of cognitive activities driven by different processes. In two experiments, we challenged this view by examining the relationship between recognition memory and inductive reasoning involving multiple forms of similarity. A common study set (members of a conjunctive category) was followed by a test set containing old and new category members, as well as items that matched the study set on only one dimension. The study and test sets were presented under recognition or induction instructions. In Experiments 1 and 2, the inductive property being generalized was varied in order to direct attention to different dimensions of similarity. When there was no time pressure on decisions, patterns of positive responding were strongly affected by property type, indicating that different types of similarity were driving recognition and induction. By comparison, speeded judgments showed weaker property effects and could be explained by generalization based on overall similarity. An exemplar model, GEN-EX (GENeralization from EXamples), could account for both the induction and recognition data. These findings show that induction and recognition share core component processes, even when the tasks involve flexible forms of similarity.

Keywords Inductive reasoning · Categorization · Recognition memory · Concepts · Computational modeling

B. K. Hayes (✉)
School of Psychology, University of New South Wales,
Kensington, NSW 2052, Australia
e-mail: b.hayes@unsw.edu.au

E. Heit
School of Social Sciences, Humanities and Arts,
University of California, Merced, CA, USA
e-mail: eheit@ucmerced.edu

In *Principles of Psychology* (1890), William James identified memory and reasoning as fundamental aspects of cognition. However, he treated them as separate components, with the coverage of the two topics appearing six chapters apart. Over a century later, cognitive psychology textbooks still treat memory and reasoning in separate chapters, six chapters apart on average (Heit & Hayes, 2008). More formally, psychological models of memory (e.g., Hintzman, 1988; Shiffrin & Steyvers, 1997) have generally focused on the retrieval of past events rather than on how such events can be used to make predictions about the future. Likewise, models of inductive reasoning (e.g., Kemp & Tenenbaum, 2009; Osherson, Smith, Wilkie, & López, 1990; Sloman, 1993) have paid little attention to the role of memory.

Some have previously attempted to bring the study of reasoning and memory closer together. Global memory models have been used to predict how people both make probabilistic judgments (e.g., Dougherty, Gettys, & Ogden, 1999) and abstract schema information from exemplars (e.g., Hintzman, 1986). Brainerd and Reyna's (1993, 2010) fuzzy trace theory has examined the overlap between memory and deductive reasoning, but with an emphasis on the functional dissociations between judgments made in each task.

We have argued for an even closer relationship between memory and reasoning (Hayes, Fritz, & Heit, *in press*; Heit & Hayes, 2005, 2011; Heit, Rotello, & Hayes, 2012). In particular, there are close parallels between the ways that people use previously studied instances to decide whether a novel probe has been seen before (i.e., recognition) and whether the properties of familiar instances generalize to the probe (i.e., inductive reasoning). In each case, probe presentation is likely to cue the retrieval of a sample of previously experienced instances and to involve a similarity comparison between the probe and sample. Where the tasks are likely to differ is in their thresholds for responding. A positive recognition response is likely to require a high level

of similarity between the study and test items. In contrast, an inductive inference that a property generalizes from a familiar base to a test item is likely to require a lower similarity threshold.

This account has been supported by recent experimentation and modeling. Heit and Hayes (2011) presented study instances under recognition (“memorize these items”) or induction (“all of these items have ‘beta cells’”) instructions. Recognition (respond “yes” if the item was seen during study) or induction (respond “yes” if the item has beta cells) judgments were then made about the items from a common test set. The rate of positive responding to novel items was higher in induction than in recognition. Nevertheless, a remarkably close relationship was found between the response rates in induction and recognition (the mean itemwise correlation between the test responses in the two tasks was .86). Moreover, both recognition and induction data could be fitted by a single modeling architecture, GEN-EX (GENERALization from EXamples), derived from exemplar models of categorization (Nosofsky, 1986, 1988).

The relationship between recognition and more complex forms of induction

These findings raise important issues regarding the relationship between memory and reasoning, suggesting that more than an analogy exists between the tasks. It may be that “recognition” and “induction” are just convenient task descriptions that refer to cognitive activities with many common processes (Heit et al., 2012). According to this “deep-correspondence” view, both tasks depend on a set of lower-level mechanisms, such as generalization (e.g., Shepard, 1987) and recollection (e.g., Rotello & Heit, 1999, 2000). Although there may be parametric differences between recognition and induction the same fundamental mechanisms are assumed to underlie both kinds of judgments (see Sun & Hélie, 2012, for a related argument).

An alternate view is that memory may be closely related to some types of induction, but not to others. Heit and Hayes (2011) examined the inductive projection of a relatively unfamiliar or “blank” property (“has beta cells”) between instances of a single study category (large dogs) and closely related probes (small, medium, and novel large dogs). Most models of induction (e.g., Osherson et al., 1990; Sloman, 1993) have assumed that the generalization of unfamiliar properties is driven by an assessment of the “overall similarity,” or amount of feature overlap, between the inductive base and target. This closely parallels the processes that are typically assumed to underlie recognition (Hintzman, 1988; Jones & Heit, 1993; Ratcliff, 1990). Hence, it may not seem all that surprising that Heit and Hayes (2011) found that the overall similarity between old and novel exemplars was important in both tasks.

In the present experiments, however, we examined the relationship between recognition and a different form of induction, one in which participants are more familiar with the property being generalized. An important finding in the induction literature is that the use of familiar or meaningful properties can change the way that the similarity between base and target items is computed in induction (e.g., Heit & Rubinstein, 1994; Medin, Coley, Storms, & Hayes, 2003; Shafto, Coley, & Baldwin, 2007). Heit and Rubinstein, for example, found that people relied on taxonomic similarity when generalizing anatomical properties across animal triads (e.g., generalization was stronger from sparrows to hawks than from tigers to hawks), but relied on ecological similarity when generalizing behavioral properties across the same triads (e.g., generalization was stronger from tigers to hawks than from sparrows to hawks).

Some researchers (e.g., Feeney, Shafto, & Dunning, 2007; Smith, Shafir, & Osherson, 1993) have suggested that such selective induction involves qualitatively different mechanisms than when the induction concerns unfamiliar properties. Whereas the latter case may be explained by overall similarity, selective induction involves knowledge of category–property relations (e.g., how predatory goals affect behavior). For example, Smith et al. introduced a new model for reasoning about meaningful properties that differed greatly from an earlier model for blank properties (Osherson et al., 1990). Although those studies did not focus on recognition, one implication of this “differentiated induction” approach is that the relationship between recognition and induction should be weaker when the induction involves familiar properties. In other words, some forms of induction potentially involve processes that differ fundamentally from those involved in recognition. For example, whereas generalization of a relatively unfamiliar property like “has beta cells” may involve overall similarity, generalization of a more familiar property like “has water-resistant skin” may involve a rule-based judgment about whether an instance is likely to live in an aquatic habitat.

The main aim of these experiments was to examine the deep-correspondence and differentiated accounts by comparing recognition and induction judgments involving familiar properties. In two experiments, we tested the recognition–induction relationship when induction involved meaningful properties that drew attention to different dimensions of similarity (i.e., selective induction). Crucially, we did not simply test the hypothesis that recognition and induction are correlated. Rather, our goal was to examine how the relationship between recognition and induction is affected by the use of meaningful properties in induction, and whether an exemplar-based approach could still account for performance in both tasks.

To this end, in each experiment we examined whether an extended version of the GEN-EX model could accommodate

inductive and recognition judgments. To preface our general approach, we retained the original GEN-EX architecture, in which recognition or inductive judgments about property generalization are based on a common process of comparison of the total similarity of a test probe to the previously experienced category members. The model was extended, however, by allowing the similarity between test probes and old category members to be computed in multiple ways (e.g., overall similarity and similarity with regard to habitat). This extension of GEN-EX is important, because it is the first time that an exemplar model has been used to explain induction and recognition involving flexible similarity relations that vary according to the property being generalized.

The main modeling comparison was between the extended GEN-EX model, with flexibility similarity, and the original GEN-EX model, with only one kind of similarity, to assess whether flexible similarity itself leads to an improved account of induction with meaningful properties. That is, we did not aim to assess the GEN-EX model in absolute terms, but rather we compared different versions of this model. To anticipate an important result, for different experimental conditions (e.g., generalizing familiar or unfamiliar properties in induction), the data supported different versions of the model.

Experiment 1

In this experiment, a common set of study and test items was presented under either recognition or induction conditions. The study items were members of a conjunctive category (aquatic mammals), and the test items belonged to the same conjunctive category or to categories that matched the study set on a single dimension (i.e., aquatic nonmammals or land mammals). The test list also included foils that differed from the study items on both dimensions (e.g., insects).

In the recognition condition, participants memorized study items. In the induction condition, they learned about animals that shared a novel property. Crucially, this property was varied across induction conditions with the aim of suggesting different kinds of similarity relations. In the induction habitat condition, study items shared a property designed to strengthen the similarity between items along the habitat dimension. In the induction mammal condition, the property was designed to strengthen mammalian similarity. In the induction blank condition, the study items shared an unfamiliar property (“has property X”).

Following Heit and Hayes (2011), we expected that induction instructions would lead to higher rates of positive responding to novel test items than would recognition instructions. In the induction conditions, however, the pattern of test responding across novel items should vary according to the property being generalized. In the induction

blank condition, high rates of responding should be limited to novel items with high overall similarity to the study set (i.e., new aquatic mammals). The two conditions involving familiar properties, however, should show evidence of selective induction. In the habitat property condition, there should be a relatively high rate of positive responding to novel aquatic mammals, whereas the reproductive property condition should show a high rate of responding to novel land mammals.

The key theoretical question was whether recognition shares core component processes with induction that involve more complex and flexible forms of similarity. In particular, we examined whether the exemplar-based model GEN-EX could be extended to account for both recognition and induction involving flexible similarity relations. The deep-correspondence approach assumes continuity between elementary and more-complex forms of induction. Namely, it assumes that inductive predictions always involve consideration of total old–new similarity, but that the basis for computing similarity between study and test items depends on the nature of the property being generalized. Following this logic, an exemplar-based account like GEN-EX may still offer a viable account of induction with familiar properties, as long as the similarity between old and new items is assessed in an appropriate way. The main modeling comparison was between the extended version of GEN-EX, with flexible similarity, versus the original version of GEN-EX, to assess whether flexible similarity might improve the fit of a unified model of induction and recognition.

Method

Participants A group of 140 university undergraduates participated for course credit. Equal numbers were randomly assigned to four conditions: recognition memory, induction blank, induction habitat, or induction mammal. One of the recognition participants responded “yes” to every foil and was replaced. A further 180 undergraduates made pairwise similarity ratings of the study and test stimuli but did not participate in the main study.

Materials The stimuli were color pictures of animals sourced from the Internet.¹ Each picture showed an animal in a left-facing side view on a white background. The study list consisted of ten pictures of aquatic mammals (e.g., gray whale, sperm whale, bottlenose dolphin, beaver, gray seal, sea lion, and walrus). The test list consisted of 50 animal pictures, including ten old items (the aquatic mammals originally studied), ten new aquatic mammals (e.g., humpback whale, orca, harbor porpoise, platypus, and leopard seal), ten new aquatic nonmammals (e.g., clownfish,

¹ We thank John Coley for providing some of the animal images.

octopus, swordfish, crab, and sea turtle), ten new land mammals (e.g., cheetah, chimpanzee, gorilla, moose, and squirrel), and ten foils (e.g., bee, eagle, butterfly, goanna, and spider).

Procedure The participants were tested individually on a desktop computer. In the recognition condition, they were instructed to memorize members of the study set. The ten study pictures were shown in random order. Each 10-cm² picture was presented for 1.5 s, with a 0.5-s interstimulus interval during which the screen was blank. A 60-s unfilled retention interval was then followed by the test phase, in which 50 test pictures were shown in random order. The participants indicated whether or not they had seen each picture by clicking the mouse on a “yes” or “no” onscreen button. The test items remained visible until a response was made.²

In the induction conditions, we used an identical procedure but had different study and test instructions. In the induction blank condition, participants were told that the goal was to learn which animals had “property X,” described as “a newly discovered property.” At test, they were told to respond “yes” to any animal that was likely to have property X. In the induction habitat condition, the goal was to learn which animals had “enzyme X,” described as “a newly discovered enzyme that assists in the development of water-resistant body covering.” The induction mammal condition had similar study instructions, except that “enzyme X” was described as “an enzyme in breast milk that is passed to the young during feeding.” In the latter two conditions, the test instruction was to respond “yes” to any item that was likely to have enzyme X.

Results

Positive test responses Table 1 shows the mean proportions of positive responses for each type of test item. These results are also shown for individual items on the left side of Fig. 1. (Note that within each type of test item, individual items are ordered arbitrarily.) Positive responses to each type of test item were entered into a series of one-way analyses of variance (ANOVAs) with four experimental groups. Three planned contrasts compared test responding in the recognition and induction conditions, responding in the induction blank and the two “nonblank” conditions, and responding in the induction habitat and induction mammal conditions.

Recognition performance was generally good, with a high hit rate on old items (old aquatic mammals), a modest false alarm rate of .25 on new aquatic mammals, and almost no false alarms to other stimulus types. The ANOVAs revealed that the rate of positive responding was higher for induction than for recognition for all stimulus types except foils [old

items, $F(1, 136) = 11.09, p < .001$, partial $\eta^2 = .08$; new aquatic mammals, $F(1, 136) = 81.15, p < .001$, partial $\eta^2 = .37$; new aquatic nonmammals, $F(1, 136) = 24.4, p < .001$, partial $\eta^2 = .15$; new land mammals, $F(1, 136) = 5.73, p = .018$, partial $\eta^2 = .04$].

Positive responding to the new aquatic-mammal items was equally high across all three induction conditions ($F < 1.0$). We found fewer positive responses to aquatic nonmammals in the blank than in the habitat condition, $F(1, 136) = 4.63, p = .03$, partial $\eta^2 = .03$. Likewise, there were fewer positive responses to land mammals in the blank than in the mammal condition, $F(1, 136) = 9.24, p = .003$, partial $\eta^2 = .06$. As expected, the patterns of positive responding differed across the induction conditions with meaningful properties. For aquatic nonmammals, rates of responding were higher when habitat rather than mammalian properties were generalized, $F(1, 136) = 38.19, p < .001$, partial $\eta^2 = .22$. This pattern reversed for responses to new land mammals, $F(1, 136) = 9.56, p = .002$, partial $\eta^2 = .07$. Hence, the property manipulation was successful in producing selective induction.

Correlations between reasoning and memory The proportions of positive responses for each of the 50 test items were averaged across participants within each experimental group, and itemwise correlations between the mean item responses in different groups were computed. Table 2 shows a strong positive relationship between “yes” responses in recognition and in all three induction conditions, replicating the findings of Heit and Hayes (2011). Positive responding in the induction blank and habitat conditions was more strongly correlated than responding in the blank and reproductive property conditions, $z = 3.25, p = .001$.

Modeling

We compared different nested versions of GEN-EX that incorporated three different kinds of judgments about the similarity between study and test items: judgments based on overall similarity, on similarity with respect to living in water, and on similarity with respect to reproduction. Inferences about a property related to water resistance could depend heavily on similarity with respect to the ability to live in water, whereas inferences about a property related to breast feeding could depend heavily on similarity with respect to reproduction. Hence, four nested versions of GEN-EX were compared. The most restricted version of the model, GEN-EX o, included overall similarity alone. This model corresponds to the original model used by Heit and Hayes (2011), and hence was the baseline for assessing other models. GEN-EX o+w+r was the most general version, including three kinds of similarity information. We also considered GENEX o+w and GENEX o+r, the intermediate models including two kinds of similarity.

² We did, however, measure response latencies. Across conditions, the mean latency for positive responses was 1.9 s.

Table 1 Experiment 1: Mean proportions (and standard errors) of positive responses and GEN-EX model predictions

Condition	Test Item Type				
	Old	New Aquatic Mammals	New Aquatic Nonmammals	New Land Mammals	Foils
Empirical Results					
Recognition	.87 (.02)	.25 (.04)	.01 (.01)	.00 (.00)	.00 (.00)
Induction blank	.94 (.02)	.72 (.05)	.32 (.06)	.05 (.03)	.03 (.02)
Induction habitat	.98 (.01)	.77 (.04)	.46 (.07)	.05 (.02)	.07 (.02)
Induction mammal	.91 (.02)	.69 (.05)	.06 (.02)	.22 (.06)	.03 (.01)
GEN-EX Model Predictions					
Recognition	.89	.23	.00	.00	.00
Induction blank	.83	.78	.34	.04	.01
Induction habitat	.86	.82	.50	.06	.01
Induction mammal	.84	.78	.06	.19	.00

Empirical similarity ratings First, we collected similarity ratings for pairs of study and test stimuli. The ten study items were paired with the 50 test items to generate 500 study–test pairs. These were divided into two sets of 250 by randomly assigning equal numbers of the following pairs to each set: old–study, old–new aquatic mammal, old–new aquatic nonmammal, old–new land nonmammal, and old–foil. Three groups of 60 participants rated 250 pairs on a seven-point scale (1 = *not similar at all*, 7 = *highly similar*), with equal numbers in each group rating pairs from one of the two stimulus sets. One group rated the pairs on “overall similarity,” a second rated pairs on similarity “with respect to their ability to live in water,” and a third rated pairs on similarity “with respect to how they reproduce.”

Table 3 shows the means for each group. Each cell is a mean over ten pairs (e.g., the study–new aquatic mammals

cell gives the mean similarity of the ten aquatic mammal test items to the study items). Old and new aquatic mammals were rated as being similar to study items for all three kinds of judgments. Foils were rated as having low similarity to the study items. Notably, aquatic nonmammals and the study items were rated as being similar with respect to habitat, but not with respect to the other dimensions. Land mammals were seen as being moderately similar to the study items with respect to reproduction, but not on the other dimensions. Hence, the three similarity judgments indexed different types of relations between the study and test items.

Model fitting The GEN-EX model is described in detail in the Appendix. Higher values of the *c* parameter indicate greater discrimination (and less generalization). Heit and Hayes (2011) found higher values of *c* for recognition than

Fig. 1 Experiment 1. Probability of a positive response to each type of test item: Empirical results and GEN-EX model predictions. Note that within each type of stimulus, each bar corresponds to an individual test item, and the individual items are ordered arbitrarily. OLD, old items; AQM, new aquatic mammals; AQNM, new aquatic nonmammals; LAM, new land mammals; LANM, land nonmammal foils

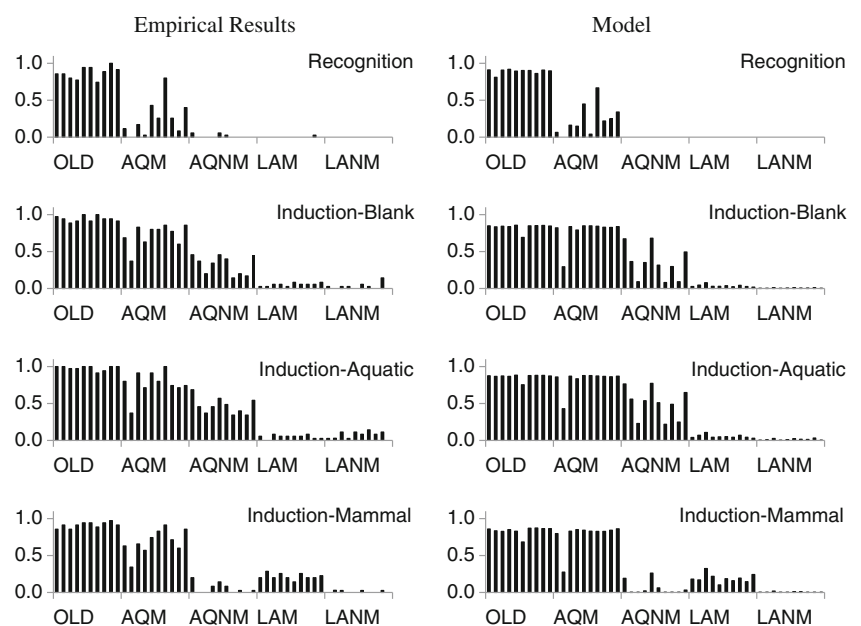


Table 2 Experiment 1: Itemwise correlations for positive responding at test

Condition	Induction Blank	Induction Aquatic	Induction Mammal
Recognition	.85*	.82*	.89*
Induction blank		.98*	.92*
Induction habitat			.87*

* Significant at $p < .001$ or better

for induction. The response bias parameter β did not play a major role in explaining the results. The most general version of the model incorporates all three types of similarity ratings, with w_o , w_w , and w_r parameters to reflect the respective weights placed on distance information derived from overall similarity, from similarity with respect to living in water, and from similarity with respect to reproduction. These were proportions that were constrained to add to 1, so in effect there were only two free parameters for weighting.

We applied the model simultaneously to the four experimental conditions, estimating separate c and β parameters, as well as w_o , w_w , and w_r parameters, for each. Thus, the most general model, GEN-EX o+w+r, incorporated three kinds of similarity, with 16 free parameters to account for 200 data points.

Table 4 shows the best-fitting parameter estimates for GEN-EX o+w+r. Overall, the model fit the data well. The correlation between the model predictions and response proportions across 200 data points was .9729, with a root-mean squared error (RMSE) of prediction of .0863. The estimated c parameters are informative. This parameter is much higher for recognition than for induction, corresponding to a high level of discrimination between old and new items. The estimated w parameters are also informative. All three kinds of similarity appear to play roles in recognition (although we would have expected a greater role for overall similarity). For the induction habitat and blank conditions, judgments seem to mainly depend on overall similarity. Notably, this pattern is different for the induction mammal condition, in which judgments depend entirely on reproductive similarity. The β parameters are all in a similar range, and we do not interpret them here.

With these parameter estimates, the predictions of the GEN-EX o+w+r model are shown in the bottom half of Table 1 and the right side of Fig. 1. The key results were captured by the model predictions: There is greater overall sensitivity to the difference between old and new items for recognition than for induction. Recognition only shows false alarms to other aquatic mammals. For induction blank and induction habitat, the patterns of generalization are similar: more positive responding to aquatic nonmammals than to land mammals. For induction mammal, this pattern reverses.

The more restricted versions of the model—GEN-EX o+w, GEN-EX o+r, and GEN-EX o—were also fit, to see whether all three kinds of similarity were required.³ Table 4 shows that the GEN-EX o+w+r model gave a significantly better fit than did any of the more restricted models, after adjustments for model complexity. In other words, including three kinds of similarity led to a better account of the data than did including just overall similarity or overall similarity plus one other kind of similarity.

Discussion

In this experiment, we examined the relationship between recognition and different types of induction judgments for a common study and test set. Overall, induction instructions led to a higher rate of positive responding to novel test items than did recognition instructions. Nevertheless, the probability of generalizing a property to test items was positively correlated with the probability of recognizing those items. Both findings are consistent with previous work (e.g., Heit & Hayes, 2011; Hayes et al., *in press*).

The most crucial novel results involved induction conditions using familiar properties. These properties were selectively generalized in the induction habitat and induction mammal conditions. The reproductive property was generalized to novel land animals, but not to nonmammalian aquatic animals. The habitat property showed the opposite generalization pattern.

These results are broadly similar to previous reports of selective induction driven by different types of familiar properties (e.g., Heit & Rubinstein, 1994; Shafto et al., 2007). However, our methods for examining inductive selectivity differed markedly from those used in previous studies. The previous studies of property effects in induction involved evaluation of the strength of written arguments. Our findings extend these effects to a task involving visual presentation of exemplars and binary decisions about property generalization. Likewise, in those previous studies, computational models were not applied.

The selective generalization pattern found in the induction habitat and induction mammal conditions was a departure from the response patterns found in recognition (as well as from the induction blank condition). Although the correlations between responding in the two tasks remained

³ As in Heit and Hayes (2011), the nested models were compared using the technique of Borowiak (1989). In brief, when model A is a nonlinear model with a free parameters estimated using a least-squares criterion, and B is a restricted version of this model with b free parameters, the likelihood ratio statistic is $\lambda = (\text{RSS A}/\text{RSS B})^{(k/2)}$, where RSS is the residual sum of squares of the model and k is the number of data points to be predicted (here, 200). Borowiak showed that $-2 \ln(\lambda)$ has a χ^2 distribution with $(a - b)$ degrees of freedom (see Rotello & Heit, 1999, 2000, and Jaeger, Cox, & Dobbins, 2012, for other applications of this technique).

Table 3 Mean similarity ratings for each type of study–test pair

Similarity Condition	Study–Test Pair				
	Study–Old	Study–New Aquatic Mammals	Study–New Aquatic Nonmammals	Study–New Land Mammals	Study–Folds
Overall similarity	4.80	4.65	3.14	1.14	1.45
Similarity with respect to living in water	6.05	5.88	5.56	1.42	1.52
Similarity with respect to reproduction	5.23	5.27	2.50	3.81	1.78

The maximum similarity rating was 7

Table 4 Experiments 1–2: Parameter estimates based on nested versions of GEN-EX

	Experiment 1	Experiment 2	
		Slow	Fast
Recognition			
β	.07	.14	.18
c	130.87	78.74	38.6
w_o	.09	.24	.95
w_w	.51	.44	.05
w_r	.49	.32	.00
Induction Blank			
β	.57		
c	7.83		
w_o	.91		
w_w	.00		
w_r	.09		
Induction Habitat			
β	.52	.26	.52
c	6.89	6.07	11.73
w_o	.85	.77	.62
w_w	.10	.23	.38
w_r	.05	.00	.00
Induction Mammal			
β	.46	.46	.24
c	15.70	12.38	6.89
w_o	.00	.10	.85
w_w	.00	.00	.10
w_r	1.00	.90	.05
RMSE	.0863	.0751	.0941
Correlation	.9744	.9821	.9729
χ^2 over o+w model	45.09***	118.96***	0.04
χ^2 over o+r model	44.11***	0.00	1.79
χ^2 over o model	45.72***	78.36***	1.79

The table shows estimated parameter values for the o+w+r model, the goodness of fit (root-mean squared error [RMSE] and correlation) for the GEN-EX o+w+r model, and comparative tests showing improvement in goodness of fit over the o+w, o+r, and o models. *** $p < .001$

strong, the induction response patterns suggested the use of more complex forms of similarity than was the case in recognition. This interpretation was supported by the results of fitting nested versions of GEN-EX. To obtain the best fit of GEN-EX to recognition and induction involving meaningful properties, we had to consider multiple forms of similarity.

On the one hand, therefore, we found that the use of familiar properties leads to some divergence between patterns of responding in induction and recognition, and that GEN-EX needs to incorporate multiple forms of similarity to deal with these data. On the other hand, the best model fits for induction and recognition always included a component of overall similarity. Most notably, a single exemplar-based architecture was able to capture the data from both complex induction and recognition tasks. On balance, we see these results as providing support for the deep-correspondence view. In Experiment 2, we aimed to provide a stronger test of this approach.

Experiment 2

Experiment 1 showed that more complex forms of similarity need to be considered when examining the relationship between recognition and induction with meaningful properties. In a new experiment, we sought to refine this conclusion by examining task conditions that promote or inhibit the use of more complex relations in induction. Previous work has shown that the use of complex forms of similarity in object comparisons and property induction is influenced by the time that is available for test decisions (Goldstone & Medin, 1994; Shafto et al., 2007). Shafto et al., for example, found that more complex forms of similarity (e.g., ecological relations between animals) were more likely to be used in induction when participants were forced to spend time deliberating about their judgments. Conversely, inductive judgments made under time pressure were based primarily on taxonomic similarity.

The aim of Experiment 2 was not so much to demonstrate that time pressure has an effect on reasoning, but to compare

the effects of time pressure on the relationship between reasoning and memory and to use these results as a further test of the modified GEN-EX model. If flexible similarity in induction is facilitated by having extra decision time, this should lead to a greater divergence between induction and recognition. Making judgments under time pressure, on the other hand, should limit the use of flexible similarity, causing patterns of induction to more closely resemble recognition judgments.

To test these predictions, we followed the Experiment 1 design but manipulated decision time. We predicted that flexible similarity would be more evident in induction when participants were given extended time for test phase decisions (slow conditions) than when test decisions were made under time pressure (fast conditions). Hence, the correlation between positive responding in induction and recognition was expected to be lower in the slow than in the fast conditions. With respect to modeling, we expected that in the slow conditions, as in Experiment 1, all three forms of similarity would need to be considered to provide an optimal fit to the recognition and induction data. For fast test judgments, however, a version of GEN-EX that incorporates fewer kinds of similarity might be sufficient. These predictions represent a further test of the deep-correspondence view, in that they suggest that a single architecture can capture changes in the induction–recognition relationship over different decision times, simply by varying the contributions of different forms of similarity to the computation of total similarity between the test and study items.

Method

Participants A group of 186 university undergraduates participated, mostly in exchange for course credit. Equal numbers were randomly allocated to the six conditions created by the factorial crossing of three types of instructions (recognition, induction habitat, and induction mammal) and decision times (fast and slow). In the fast condition, 18 participants failed to respond within the response deadline on more than 75 % of the test trials and were replaced.

Materials and procedure The study phase procedure was the same as in Experiment 1, except that, because we were primarily interested in changes in the use of flexible similarity, we omitted the induction blank condition. The test procedure was similar to Experiment 1, except for the timing manipulation.

Those in the fast condition responded “as quickly as possible without sacrificing accuracy.” Each test item was presented onscreen, together with “yes” and “no” response buttons, for 1.5 s. Participants had to respond within this period or a buzzer sounded and a “time out” message was

displayed (and the data for that trial were excluded). Note that the maximum response time in this condition was below the mean response latency observed in Experiment 1 (see note 2). In the slow condition, participants were told to “think carefully” before responding. Each test item was visible for 9 s, but no response could be made during this period (the response buttons were grayed out). After this period, the response buttons changed color and a response could be made.

Results and discussion

Positive test responses The mean positive responses at test are shown in the top panel of Table 5 and are broken down by items in the left halves of Figs. 2 and 3. Group differences between response rates were examined in a series of one-way ANOVAs. Planned contrasts compared the response rates for recognition and induction (averaged across induction conditions) and for the induction habitat and induction mammal conditions. For all stimulus types, the rate of positive responding was higher for induction than for recognition (all p s < .001).

Those generalizing aquatic properties were more likely to respond positively to new aquatic nonmammals than were those in the mammal property condition, $F(1, 180) = 134.17, p < .001$, partial $\eta^2 = .43$. As predicted, this difference was accentuated under slow decision conditions, $F(1, 180) = 14.86, p < .001$, partial $\eta^2 = .08$. The pattern reversed for responses to new land mammals, for which positive responding was more common in the mammal property than in the habitat property condition, $F(1, 180) = 20.78, p < .001$, partial $\eta^2 = .10$. In this case, responding was not affected by decision time (F s < 2.0).

Correlations between reasoning and memory responses The correlations between responses to individual test items in each condition are given in Table 6. The table shows a strong positive correlation between responding in the recognition and induction mammal conditions, and that this correlation was unaffected by decision time. Notably, the correlation between responding in the recognition and induction habitat conditions was lower in the slow than in the fast condition, $z = 1.67, p = .05$, one tailed.

Modeling

The slow and fast conditions were modeled separately. Within each decision condition, we applied the model simultaneously to the recognition and induction groups. We estimated separate c and β parameters, as well as w_o , w_w , and w_r parameters for each condition. For each condition, in the GEN-EX $o+w+r$ model, 12 free parameters were used to account for 150 data points.

Table 5 Experiment 2: Mean proportions (and standard errors) of positive responses and GEN-EX model predictions

Condition	Test Item Type				
	Old	New Aquatic Mammals	New Aquatic Nonmammals	New Land Mammals	Foils
Empirical Results					
Fast					
Recognition	.85 (.03)	.38 (.03)	.03 (.01)	.00 (.00)	.00 (.00)
Induction habitat	.99 (.01)	.87 (.02)	.48 (.07)	.05 (.02)	.06 (.02)
Induction mammal	.96 (.01)	.80 (.04)	.15 (.03)	.31 (.07)	.02 (.01)
Slow					
Recognition	.87 (.02)	.39 (.05)	.08 (.03)	.03 (.01)	.01 (.01)
Induction habitat	.99 (.01)	.92 (.03)	.77 (.06)	.08 (.04)	.08 (.02)
Induction mammal	.90 (.03)	.74 (.04)	.12 (.03)	.22 (.07)	.00 (.00)
GEN-EX Model Predictions					
Fast					
Recognition	.84	.43	.00	.00	.00
Induction habitat	.86	.82	.50	.06	.01
Induction mammal	.91	.85	.49	.00	.00
Slow					
Recognition	.85	.39	.00	.00	.00
Induction habitat	.94	.92	.77	.14	.06
Induction mammal	.85	.80	.11	.21	.01

The best-fitting parameter estimates are shown in Table 4. The model fit the data well for both slow and fast judgments. For slow judgments, the correlation between the model predictions and the response proportions was .9821, and the RMSE was .0751. The fit for fast judgments was also good, although slightly worse than the fit for slow judgments (the correlation between predictions and response proportions was .9729, with an RMSE of .0941).

In general, the estimated *c* parameters show a pattern similar to the one found in Experiment 1—namely, higher values for recognition than for induction, corresponding to a higher level of discrimination between old and new items. For the slow

conditions, the estimated *w* parameters also show a pattern similar to the one in Experiment 1: Namely, recognition is influenced by all three types of similarity, induction about habitat is influenced mainly by overall similarity and aquatic similarity, and induction about a mammalian property is influenced mainly by reproductive similarity. The fast conditions show a less specialized use of similarity: Namely, recognition shifts to being based on overall similarity, and most importantly, induction about a mammalian property shifts from reproductive similarity to also become based on overall similarity. In general, the fast conditions appear to be more sensitive to overall similarity than are the slow conditions.

Fig. 2 Experiment 2 fast condition. Probability of a positive response to each type of test item: Empirical results and GEN-EX model predictions. Note that within each type of stimulus, each bar corresponds to an individual test item, and the individual items are ordered arbitrarily. OLD, old items; AQM, new aquatic mammals; AQNM, new aquatic nonmammals; LAM, new land mammals; LANM, land nonmammal foils

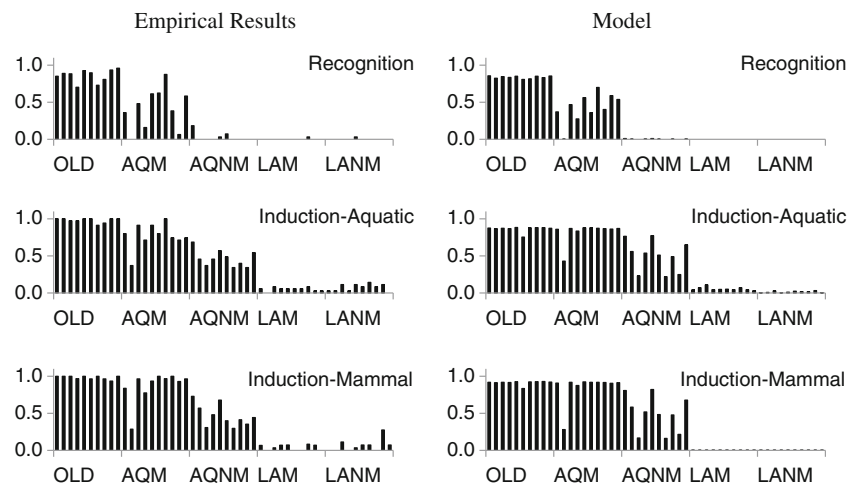
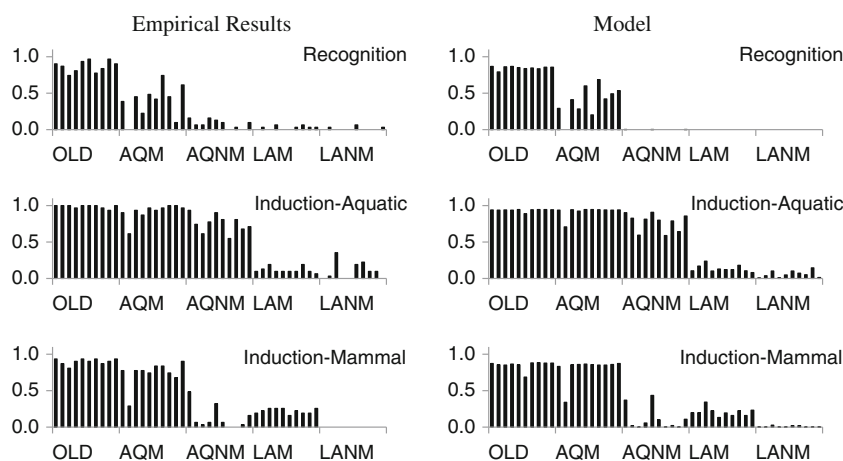


Fig. 3 Experiment 2 slow condition. Probability of a positive response to each type of test item: Empirical results and GEN-EX model predictions. Note that within each type of stimulus, each bar corresponds to an individual test item, and the individual items are ordered arbitrarily. OLD, old items; AQM, new aquatic mammals; AQNM, new aquatic nonmammals; LAM, new land mammals; LANM, land nonmammal foils



The model predictions based on these parameter estimates are shown in the lower panel of Table 5 and the right halves of Figs. 2 and 3. The key empirical results were captured. In slow conditions, there is greater sensitivity to the difference between old and new items for recognition than for induction, and recognition only shows false alarms to other aquatic mammals. The induction habitat condition shows more positive responding to aquatic nonmammals than to land mammals, and for induction with a mammal property, the pattern is reversed. In the fast conditions, there is sharper discrimination between old and new items for recognition than for induction, but little difference in the generalization patterns of the two induction conditions.

More restricted models involving overall similarity alone (GEN-EX o) or two forms of similarity (GEN-EX o+w, GEN-EX o+r) were again applied, and their fits were compared to the more general model. As is shown in Table 4, for the slow conditions, the GEN-EX o+w+r model gave a significantly better fit than did two of the three more restricted models. However, it was not significantly different in fit than the model using overall similarity plus reproductive similarity. In other words, adding reproductive similarity led to a significant improvement in goodness of fit beyond overall similarity, but in this case, adding water similarity did not improve matters further. Although this finding is somewhat different than that of Experiment 1, note that habitat similarity was estimated to have a lower

impact on induction than the two other kinds of similarity in Experiment 1. Crucially, for fast conditions, the model with three kinds of similarity was not significantly better than any of the restricted models. In other words, overall similarity was sufficient to explain speeded memory and reasoning responses.

To summarize, we again found evidence of selective use of different types of similarity in induction, but the extent of selective induction was modulated by decision time. Slowing decisions led to enhanced property effects in the induction habitat condition and further divergence between the patterns of responding in induction and recognition. Speeding up induction decisions tended to suppress the use of complex similarity relations. The whole pattern of results could be accounted for with the GEN-EX model, by incorporating only overall similarity for the fast conditions, and more specialized similarity as well for the slow conditions. On balance, we again take these results as being consistent with the deep-correspondence view, treating the difference between the original and extended versions of GEN-EX as a quantitative one in terms of additional parameter values.

General discussion

In these two experiments, we examined relations between inductive reasoning and recognition involving flexible forms of similarity. In both, we found that the use of meaningful properties led to selective inductive judgments based on different kinds of similarity. This led to some divergence between the response patterns for induction versus recognition, especially when people were encouraged to deliberate about their judgments. Nevertheless, the empirical relationship between induction and recognition test responding remained relatively strong (with a minimum correlation of .73). Moreover, an exemplar-based model that incorporated multiple similarity measures was able to account for both recognition and induction on the basis of flexible similarity.

Table 6 Experiment 2: Itemwise correlations for positive responding at test in the fast and slow decision conditions

	Cond.	Induction Habitat	Induction Mammal
Recognition	Fast	.86*	.89*
	Slow	.73*	.90*
Induction habitat	Fast		.88*
	Slow		.78*

* Significant at $p < .001$ or better

GEN-EX modeling provided additional insights into the processes that underlie induction and recognition. One finding (similar to those of Hayes et al., *in press*, and Heit & Hayes, 2011) was that speeded recognition and induction (Exp. 2) could be explained by generalization based on a single similarity measure, assuming a broader gradient of generalization for induction. A second important modeling result was that the property effects found in self-paced (Exp. 1) and slow (Exp. 2) induction could not be explained by the original version of GEN-EX. To account for these data, we extended GEN-EX to include multiple similarity relations (i.e., similarity with respect to habitat and reproduction). In general, the parameters estimated from modeling were informative—for instance, showing different generalization gradients for recognition versus induction, and different uses of similarity information for induction about different properties.

Implications for the relationship between memory and reasoning

The results have important implications for understanding the relationship between inductive reasoning and memory. According to the differentiated view, induction works differently depending on whether unfamiliar or familiar properties are being generalized (see, e.g., Smith et al., 1993). For example, whereas an unfamiliar property like “has enzyme *x*” might invite similarity-based induction, a more familiar property related to aquatic habitat might invite the use of a rule based on prior knowledge, such as respond “yes” to all aquatic creatures. Hence, it should be difficult to explain induction and recognition involving complex similarity within a single exemplar-based architecture.

By contrast, the deep-correspondence view assumes a continuity in process between induction involving overall similarity and induction involving more abstract relations. According to this view, a close relationship may still exist between the processes that underlie recognition and induction involving more abstract forms of similarity, especially when the specific requirements of each task (e.g., the response instructions) are made more comparable.

Overall, our empirical and modeling results support the deep-correspondence view. In both studies, we have found higher rates of positive responding to old test items than to similar lures in conditions involving meaningful properties, including those involving unfamiliar properties. This suggests that exemplar similarity was crucial to induction and recognition and that participants were not simply responding on the basis of rules suggested by prior knowledge of the stimuli. Indeed, further modeling, summarized in the [Appendix](#), showed that rule-based models without an overall-similarity component did a poor job of explaining the recognition and induction data.

The logic of our studies has been to take strong correlations between tasks, and the ability of the same model with only parametric changes to account for multiple tasks, as evidence for the deep-correspondence view. Hence, we have aimed to make a positive contribution by articulating the relatively novel deep-correspondence view and to offer GEN-EX as an existence proof that a single framework can capture induction and recognition judgments involving complex forms of similarity. We acknowledge what may seem to be a limitation of our work: We did not implement formal models corresponding to the differentiated-induction view. Proponents of that view might argue that the correlations would need to be even stronger, or they might propose that separate models for recognition, induction with blank properties, and induction with meaningful properties could also fit the data. We have little doubt that separate models could account for these data, albeit at a considerable cost in parsimony. Alternatively, it might be argued that the extended version of GEN-EX is fundamentally different from the original. We see this difference, however, as being quantitative rather than qualitative, involving the addition of similarity functions with nonzero parameters for sensitivity.

To some it may seem self-evident that a close relation exists between reasoning and memory, when responses on both tasks can be affected by the similarity between familiar and novel items. It is worth restating that we did not simply test the hypothesis that recognition and induction are correlated due to their common use of similarity, but rather we examined how the use of similarity varies for induction about different properties and due to response time manipulations. Additionally, it is important to note that some theoretical approaches predict that performance levels on memory and reasoning tasks will only be weakly related or, in some cases, statistically independent. For example, fuzzy trace theory (Brainerd & Reyna, 1993, 2010) suggests that reasoning and memory are often driven by different processing components (e.g., reasoning is predominantly driven by gist representations, and recognition is affected by verbatim traces). As a result, accuracy in deriving valid deductive conclusions can be stochastically independent of memory for the relevant premises (e.g., Brainerd & Reyna, 1993). Our results suggest that such memory independence may be limited to deductive reasoning. By contrast, in induction, memory for the details of the inductive base (i.e., the study instances in the present experiments) appears to be critical for property generalization (see also Heit & Hayes, 2011; Heit et al., 2012).

In general, we hope that this work will spur on further model development by other researchers. For example, recent Bayesian models of induction (e.g., Kemp & Tenenbaum, 2009) have improved on earlier models (e.g., Heit, 1998, 2000) by applying a similar framework to induction with both highly meaningful and less-meaningful

properties. However, these Bayesian models of induction have not yet been applied to recognition. Likewise, another promising approach to explain studies like our own may be to extend connectionist models (see Glick, 2011). The aim of our own modeling has been to provide an existence proof for the viability of a unified model and to show the role of flexible similarity, rather than to rule out other possible unified models of recognition and induction.

Our results suggest that “recognition” and “induction” may be useful task descriptions, but that these labels do not necessarily map onto fundamental distinctions in human cognition. If we are looking for critical distinctions between different types of cognitive activities, it may be better to focus on the distinction between judgments based on broad or narrow definitions of similarity or on fast as opposed to slow judgments (cf. Heit et al., 2012).

Relationship to previous work on the relations between memory, reasoning, and categorization

An important implication of these findings is that a single exemplar-based framework can be used to explain recognition, induction involving blank properties, and induction involving meaningful properties. Although different types of similarity were involved in the generalization of novel and familiar properties, the core assumptions of GEN-EX (i.e., generalization based on total similarity, where similarity follows a negative exponential function of the distance between test and study items) were maintained in modeling the data from all induction and recognition conditions.

Exemplar models incorporating assumptions similar to those of GEN-EX have been successful in accounting for patterns of categorization and recognition of the same stimulus sets (e.g., Nosofsky, 1986, 1988, Nosofsky, Little, & James, 2012) but have rarely been applied to inductive reasoning (see Heit, 1992; Holyoak, Lee, & Lu, 2010). Currently, an intense debate concerns the degree of overlap between the processes involved in categorization and induction (cf. Gelman, 2003; Sloutsky, 2010). Our view is that while there are commonalities between categorization and induction, differences seem to exist as well. For example, categorization directs attention to diagnostic features that discriminate between members of contrasting categories, whereas feature inference directs attention to relations between features within a category (e.g., Sweller & Hayes, 2010; Yamauchi & Markman, 1998; see Kemp & Jern, 2009, and Markman & Ross, 2003, for discussion of other differences between induction and categorization). Hence, the application of the exemplar-based GEN-EX model to more complex forms of induction and

recognition represents a significant extension to previous work with exemplar models.

Another important finding was that the time available for an inductive judgment influenced the types of similarity relations used in induction and recognition. Few previous studies have investigated the time course of processing in induction. By contrast, considerable attention has been devoted to time-dependent changes in recognition processes (e.g., Gronlund & Ratcliff, 1989; Rotello & Heit, 1999, 2000) and categorization (e.g., Lamberts, 2002; Little, Nosofsky, & Denton, 2011). Gronlund and Ratcliff, for example, showed that the details of studied items can influence recognition under short decision deadlines, but that judgments based on relations between items require more decision time. The degree to which multiple stimulus dimensions need to be considered in order to accurately categorize stimuli has also been shown to follow a predictable time course (Little et al., 2011).

This suggests that future research on inductive reasoning would benefit from a fine-grained analysis of changes in the availability of different kinds of stimulus information over the course of induction. This would be particularly interesting to examine in paradigms similar to those used here, where multiple forms of similarity can contribute to inductive judgments.

An important theme of the present study was that similarity can be treated as a flexible construct in memory and reasoning. Of course, additional processes may influence responding in each task. For example, some forms of memory (e.g., the recall-to-reject phenomenon examined by Rotello & Heit, 1999) may operate like logical reasoning. It may not be possible to reduce such effects to a weighted combination of different types of similarity. Future work that examines such forms of reasoning and memory may uncover more dramatic dissociations than those found here. Nevertheless, the present work shows that a much closer relationship exists between recognition and induction than has previously been acknowledged. Recognition and induction appear to share component processes, even when these tasks involve flexible forms of similarity.

Author note This work was supported by Australian Research Council Discovery Grant No. DP1093197 to both authors, and by National Science Foundation Grant No. BCS-0616979 to the second author. We thank Elizabeth Balboa, Melony Bowling, Wendy Contreras, Graham Ellis, Ryan Hoffman, Melissa Lim, Helen Paton, Ann Martin, Alex Parnell, and Aljane Whitaker for their assistance with this research.

Appendix

The original version of GEN-EX is closely related to Nosofsky's (1986) generalized context model (GCM) of

categorization, which was itself, as its name implies, a generalization of Medin and Schaffer’s (1978) context model (but see Heit & Hayes, 2011, pp. 78–79, for a discussion of differences between GEN-EX and GCM). GEN-EX is embodied by two equations. Equation A1 shows the familiarity rule: The familiarity of each test stimulus, *fam*, equals its summed similarity to *n* studied items. Similarity is assumed to be a negative exponential function of the distance, *dist*, between the test and study items. The free parameter *c* reflects the specificity of responding to test items; lower values of *c* correspond to broader generalization, while higher values correspond to narrower generalization gradients.

$$fam(test) = \sum_{i=1}^n \exp[-c \text{dist}(test, \text{study}_i)] \tag{A1}$$

$$resp(test) = \left\{ \frac{fam(test)}{fam(test) + \beta} \right\} \tag{A2}$$

$$fam(test) = \sum_{i=1}^n \exp\{-c[w_o \text{dist}_o(test, \text{study}_i) + w_w \text{dist}_w(test, \text{study}_i) + w_r \text{dist}_r(test, \text{study}_i)]\} \tag{A3}$$

The three *w* parameters are weighting parameters; they represent the relative contributions of each source of information. These parameters were constrained to add to 1. The three components of total distance, *dist_o*, *dist_w*, and *dist_r*, were derived, respectively, from the three types of similarity judgments: overall similarity, similarity with respect to ability to live in water, and similarity with respect to reproduction. Again, it was assumed that similarity would be a negative exponential function of psychological distance (e.g., Nosofsky, 1986; Shepard, 1987), as is illustrated by Eq. A4.

$$sim(x, y) = \exp[-c \text{dist}(x, y)] \tag{A4}$$

Similarity ratings were normalized to the range of 0 to 1. Equation A4 was used to convert similarity ratings to distances; that is, Eq. A4 was solved for distance as a logarithmic function of similarity, to calculate distances as a function of similarity. For this purpose, the *c* parameter was arbitrarily set at 1 (allowing *c* to vary did not improve model fit).

Rule-based extensions

The original Heit and Hayes (2011) article also considered the possibility of rule-based responding—for instance, “respond yes to all large dogs.” There, it was found that

The response rule is shown in Eq. A2. Essentially, the probability of a positive response is a monotonic function of a test item’s familiarity. The response rule has a single scaling parameter, *β*. A lower value of *β* corresponds to a greater overall tendency to respond positively.

This modeling process was carried out separately for recognition judgments and for each type of induction judgment, with each type of judgment having its own *c* and *β* parameters. Hence, it was possible, for example, to compare the *c* parameter for recognition versus induction, to assess the idea that recognition has narrower generalization and that induction has broader generalization.

For each type of judgment, we treated overall distance as a linear combination of three sources of information, derived from the empirical ratings of overall similarity, similarity with respect to ability to live in water, and similarity with respect to reproduction. Hence, Eq. A1 could be rewritten as Eq. A3.

including a rule-based component generally improved the fit of GEN-EX, although exemplar similarity dominated the responses. Rule-based responding is not the focus of the present article, where we simply aimed to show that a model based on flexible exemplar similarity is adequate to explain results from memory and reasoning. In general, here we came to conclusions comparable from modeling. For example, in Experiment 1, we implemented rule-based components to respond positively to all mammals, all aquatic creatures, and all aquatic mammals. Adding these rules did significantly improve the overall fit of the model, beyond the role of exemplar similarity, although, again, the estimated role of exemplar similarity was much higher than the role of rule-based responding. A model based on rule-based responding alone fit the data relatively poorly, and adding exemplar similarity to that model improved the fit very substantially, suggesting a critical role for similarity between the study and test items. Overall, the need for exemplar similarity is evident from the close correspondence between data and the model on individual items in Figs. 1, 2, and 3. In particular, a rule-based model does not explain the higher rates of responding to old items in the figures, as compared with new aquatic mammals (AQM).

References

- Borowiak, D. S. (1989). *Model discrimination for nonlinear regression models*. New York, NY: Marcel Dekker.
- Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review*, *100*, 42–67. doi:10.1037/0033-295X.100.1.42
- Brainerd, C. J., & Reyna, V. F. (2010). Recollective and nonrecollective recall. *Journal of Memory and Language*, *63*, 425–445. doi:10.1016/j.jml.2010.05.002
- Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review*, *106*, 180–209. doi:10.1037/0033-295X.106.1.180
- Feeney, A., Shafto, P., & Dunning, D. (2007). Who is susceptible to conjunction fallacies in category based induction? *Psychonomic Bulletin and Review*, *14*, 884–889. doi:10.3758/BF03194116
- Gelman, S. A. (2003). *The essential child*. New York, NY: Oxford University Press.
- Glick, J. J. (2011). *Uncovering the organization of semantic structure with similarity and inductions*. Unpublished doctoral dissertation, Stanford University.
- Goldstone, R. L., & Medin, D. L. (1994). Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 29–50. doi:10.1037/0278-7393.20.1.29
- Gronlund, S. D., & Ratcliff, R. (1989). The time-course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 846–858. doi:10.1037/0278-7393.15.5.846
- Hayes, B. K., Fritz, K., & Heit, E. (in press). The relationship between memory and inductive reasoning: Does it develop? *Developmental Psychology*. doi:10.1037/a0028891
- Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*, 341–380. doi:10.1016/0010-0285(92)90011-P
- Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, UK: Oxford University Press.
- Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*, 569–592. doi:10.3758/BF03212996
- Heit, E., & Hayes, B. K. (2005). Relations between categorization, induction, recognition and similarity. *Journal of Experimental Psychology: General*, *134*, 596–605. doi:10.1037/0096-3445.134.4.596
- Heit, E., & Hayes, B. K. (2008). Predicting reasoning from visual memory. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 83–88). Austin, TX: Cognitive Science Society.
- Heit, E., & Hayes, B. K. (2011). Predicting reasoning from memory. *Journal of Experimental Psychology: General*, *140*, 76–101. doi:10.1037/a0021488
- Heit, E., Rotello, C. M., & Hayes, B. K. (2012). Relations between memory and reasoning. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 57–101). San Diego, CA: Academic Press.
- Heit, E., & Rubinstein, J. (1994). Similarity and property effects in inductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 411–422. doi:10.1037/0278-7393.20.2.411
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411–428. doi:10.1037/0033-295X.93.4.411
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528–555. doi:10.1037/0033-295X.95.4.528
- Holyoak, K. J., Lee, H. S., & Lu, H. (2010). Analogical and category-based inference: A theoretical integration with Bayesian causal models. *Journal of Experimental Psychology: General*, *139*, 702–727. doi:10.1037/a0020488
- Jaeger, A., Cox, J. C., & Dobbins, I. G. (2012). Recognition confidence under violated and confirmed memory expectations. *Journal of Experimental Psychology: General*, *141*, 282–301. doi:10.1037/a0025687
- James, W. (1890/1983). *Principles of psychology*. Cambridge, MA: Harvard University Press.
- Jones, C. M., & Heit, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 799–812. doi:10.1037/0278-7393.19.4.799
- Kemp, C., & Jern, A. (2009). A taxonomy of inductive problems. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 255–260). Austin, TX: Cognitive Science Society.
- Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58. doi:10.1037/a0014282
- Lamberts, K. (2002). Feature sampling in categorization and recognition of objects. *Quarterly Journal of Experimental Psychology*, *55A*, 141–154. doi:10.1080/02724980143000208
- Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1–27. doi:10.1037/a0021330
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592–613. doi:10.1037/0033-2909.129.4.592
- Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*, 517–532. doi:10.3758/BF03196515
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238. doi:10.1037/0033-295X.85.3.207
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57. doi:10.1037/0096-3445.115.1.39
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700–708. doi:10.1037/0278-7393.14.4.700
- Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Science*, *109*, 333–338. doi:10.1073/pnas.1111304109
- Osherson, D. N., Smith, E. E., Wilkie, O., & López, A. (1990). Category-based induction. *Psychological Review*, *97*, 185–200. doi:10.1037/0033-295X.97.2.185
- Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285–308. doi:10.1037/0033-295X.97.2.285
- Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, *40*, 432–453. doi:10.1006/jmla.1998.2623
- Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907–922. doi:10.3758/BF03209339
- Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review*, *14*, 890–894. doi:10.3758/BF03194117

- Shepard, R. N. (1987). Toward a universal Law of Generalization for psychological science. *Science*, *237*, 1317–1323. doi:[10.1126/science.3629243](https://doi.org/10.1126/science.3629243)
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*, 145–166. doi:[10.3758/BF03209391](https://doi.org/10.3758/BF03209391)
- Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231–280. doi:[10.1006/cogp.1993.1006](https://doi.org/10.1006/cogp.1993.1006)
- Sloutsky, V. S. (2010). Similarity, induction, naming and categorization: A bottom-up approach. In S. P. Johnson (Ed.), *Neoconstructivism: The new science of cognitive development* (pp. 274–294). New York, NY: Oxford University Press.
- Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition*, *49*, 67–96. doi:[10.1016/0010-0277\(93\)90036-U](https://doi.org/10.1016/0010-0277(93)90036-U)
- Sun, R., & Hélie, S. (2012). Psychologically realistic cognitive agents: Taking human cognition seriously. *Journal of Experimental and Theoretical Artificial Intelligence*, *1*–28. doi:[10.1080/0952813X.2012.661236](https://doi.org/10.1080/0952813X.2012.661236)
- Sweller, N., & Hayes, B. K. (2010). More than one kind of inference: Re-examining what's learned in feature inference and classification. *Quarterly Journal of Experimental Psychology*, *63*, 1568–1589. doi:[10.1080/17470210903438547](https://doi.org/10.1080/17470210903438547)
- Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*, 124–148. doi:[10.1006/jmla.1998.2566](https://doi.org/10.1006/jmla.1998.2566)