CHAPTER TWO

Relations Between Memory and Reasoning

Evan Heit, Caren M. Rotello *and* Brett K. Hayes

Contents

Abstract

Memory and reasoning are traditionally treated as separate issues by psychologists. In contrast, two lines of research are presented that take advantage of analogies between memory and reasoning and explore the deeper connections between them. The first line of research takes theoretical ideas (two-process accounts) and methodological tools (signal detection analysis, receiver operating characteristic curves) from memory research and applies them to important issues in reasoning research: relations between induction and deduction, and the belief bias effect. The results showed that induction and deduction can be conceived of as drawing on mixtures of two underlying reasoning processes, corresponding to heuristic and analytic processes. For belief bias, it was found that this result can be explained in terms of a simple response bias, without assuming different accuracy for reasoning about believable versus unbelievable

Psychology of Learning and Motivation, Volume 57
ISSN 0079-7421, DOI: 10.1016/B978-0-12-394293-7.00002-9

arguments. The second line of research introduces a task in which subjects make either memory or reasoning judgments for the same set of stimuli. Other than broader generalization for reasoning than memory, the results were similar for the two tasks, even across a variety of experimental manipulations. It was possible to simultaneously explain both tasks within a single framework, using either exemplar modeling or signal detection modeling.

1. Introduction

By tradition, memory and reasoning are kept apart. Each topic has most often been studied with its own experimental paradigms, largely by different researchers addressing different questions, and resulting in reasoning phenomena and memory phenomena being addressed by separate theories. There are even separate academic conferences addressing either memory or reasoning. We recently examined modern cognitive psychology textbooks (Heit & Hayes, 2008), and observed that memory and reasoning are always treated in separate chapters, six chapters apart on average. This separation has a distinguished history. Notably, in William James's (1890) *Principles of Psychology*, memory and reasoning were also covered six chapters apart from each other. Of course, there are exceptions to this generalization. For example, research on metacognition sometimes addresses how people reason about their own memories (e.g., Finn & Metcalfe, 2008; Townsend & Heit, 2011), and research has shown false memories can be created through reasoning (Brainerd & Reyna, 1993; Sloutsky & Fisher, 2004). Working memory research sometimes examines the relation between memory capacity and intelligence (e.g., Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Oberauer, Süss, Wilhelm, & Wittmann, 2008). Modeling frameworks such as Bayesian models (Chater, Oaksford, Hahn, & Heit, 2010) and connectionist models (e.g., Rogers & McClelland, 2004) have been applied to both reasoning and memory. Still, the usual assumption is that reasoning and memory are very different cognitive activities.

To spell out this point further, we briefly outline some possible views on the relations between memory and reasoning.

1. Memory and reasoning are **unrelated** or independent, in the sense that studying one tells us little or nothing about the other (e.g., Brainerd & Kingma, 1985). On this view, one would expect different representations and processes to be used in the two domains, with little to no overlap. Memory performance would not reliably predict reasoning performance within an individual, nor across stimulus sets.
2. There is a good **analogy** between the kinds of questions that researchers have asked about memory and reasoning (e.g., do they rely

on a single or multiple component processes?). This means that approaches used to study one type of task may give insights into the other (see Medin, Goldstone, & Markman, 1995, for related arguments regarding parallels between decision making and similarity judgment, and Ross, 1996, for arguments regarding parallels between categorization and problem solving).

3. Memory and reasoning are **interdependent**, e.g., memory processes may rely on reasoning processes and vice versa. Some of the work on false memory (e.g., Sloutsky & Fisher, 2004), metamemory (e.g., Finn & Metcalfe, 2008), working memory (e.g., Jaeggi et al., 2008), analogy (e.g., Kokinov & Petrov, 2001), and links between memory and imagining future events (e.g., Schacter, Addis, & Buckner, 2007) falls under this view. (See Bartek, Lewis, Vasishth, & Smith, 2011, and Vlach & Sandhofer, 2011 for related arguments that language processing and acquisition depend on memory mechanisms.)
4. Memory and reasoning are part of a **bigger whole**, whether this is a general higher-order cognitive system (Fodor, 1983) or some specific cognitive architecture (e.g., Bayesian models, Chater et al., 2010, neural networks, e.g., Rogers & McClelland, 2004, or production systems, e.g., Anderson & Lebiere, 1998). Hence they will have many commonalities.
5. Memory and reasoning depend on set of lower-level **common mechanisms**, such as generalization (e.g., Shepard, 1987) and recollection (e.g., Rotello & Heit, 1999, 2000). We note that this view may be difficult to distinguish empirically from the previous view, and the difference could be more of a reflection of a researcher's theoretical approach.
6. Memory and reasoning are the **wrong slices** of the cognitive pie. There are cognitive mechanisms such as generalization and recollection that should be studied in their own right, but there is so much overlap between memory and reasoning that the dividing line is arbitrary (Heit & Hayes, 2005).
7. It does not even make sense to speak of discrete cognitive mechanisms, because cognition is **emergent** from low-level interactions such as in dynamical systems (e.g., Barsalou, Breazeal, & Smith, 2007; Van Orden, Jansen op de Haar, & Bosman, 1997). Various cognitive capacities arise from these interactions, and describing these as discrete functions or mechanisms is an approximation at best.

Our own views are most sympathetic toward the **analogy, interdependent, bigger whole, common mechanisms**, and **wrong slices** views. Section 2 of this chapter will make a strong case for analogies between memory and reasoning, supporting the **analogy** view. In this section we show that considerable progress can be made on some long-standing questions about human reasoning by applying the logic and methods used to study memory. This work serves as a counterpoint to the **unrelated** view. Although we do not dispute the **interdependent** view, and there is ample evidence in support, it is not our focus here. Section 3 of this chapter shows a great deal of empirical similarity between memory and reasoning, supporting the **bigger whole** and **common mechanisms** views. We have developed a common model for both memory and reasoning, also supporting these views. In section 4 of this chapter, we return to the possibility that cognitive psychologists have carved up the cognitive pie incorrectly, as in the **wrong slices** view. Finally, the **emergent** view is an interesting perspective, but we will not address it directly.

We next turn to a line of work that takes advantage of analogies between memory and reasoning, borrowing experimental and analytical tools from memory research and applying them to important problems in reasoning research. For clarity, we note that we do not attempt to address all of memory in this chapter, but instead focus on recognition memory. Likewise we do not address all forms of reasoning, instead focusing on inductive reasoning and, to a lesser extent, deductive reasoning. No doubt there would be an even richer story to tell by addressing further aspects of memory and reasoning.

2. Applying Memory Methods to Reasoning

Heit and Rotello (2005) pointed out a "striking parallel" between memory and reasoning, namely that in both areas of research, there is a central debate about whether there are one or two underlying processes. In memory research there is an important distinction between one- and two-process accounts (Rotello & Heit, 1999; Yonelinas, 2002). According to two-process accounts of recognition memory, recognition judgments depend on a quick, approximate, familiarity-based process and a slower, more deterministic process based on specific item recollection. In effect, there are two different kinds of recognition, because either process could dominate or even fully determine a recognition judgment.

In contrast, according to one-process accounts, it is not necessary to assume two processes to explain experimental results. This distinction has come up in the context of whether remembering and knowing correspond to different processes. According to some researchers (Donaldson, 1996; Dougal & Rotello, 2007; Dunn, 2008; Wixted & Stretch, 2004) the distinction between remembering and knowing is simply a matter of a criterion shift, i.e., both judgments are based on a common scale of memory strength, but there is a stricter criterion for saying that something is directly remembered. Hence, in terms of signal detection theory (SDT), the difference between remembering and knowing should appear as a change in response criterion rather than sensitivity. However, other assessments (Gardiner, Ramponi, & Richardson-Klavehn, 2002; Rotello, Macmillan, & Reeder, 2004; Wixted & Mickes, 2010) have rejected a one-dimensional signal detection model. In particular, in memory research there are standard signs taken as evidence against a single process, such as unequal sensitivity for different types of judgments on the same memory probes, slope differences in receiver operating characteristic (ROC) curves, and a non-monotonic relationship between the two types of judgments across a set of probes. On this basis, Rotello et al. proposed a two-dimensional model, incorporating information from familiarity and recollection.

In reasoning research, there are likewise two groups of researchers, who have taken opposing positions on whether there are one or two kinds of reasoning. Some researchers have assumed that induction and deduction depend on the same cognitive mechanisms. Several influential research programs embody this one-process view, by applying the same modeling framework to both inductive and deductive problems, assuming a single scale of evidence for argument strength. For example, Oaksford and Chater (2007) showed that a Bayesian reasoning model, probabilistic in nature, can be successfully applied to problems of deduction. Note that Oaksford and Chater (2002) themselves have not denied that people can explicitly perform deductive tasks under limited circumstances. The key point is that this line of work offers a unifying framework for problems of deduction and induction, based on a single scale of probability. Likewise, Johnson-Laird (1994) explained how mental models theory, typically applied to problems of deduction, can also be applied to problems of induction. Osherson, Smith, Wilkie, Lopez, and Shafir (1990), as well as Sloman (1993), presented models of inductive reasoning that, without additional assumptions, account for some deductive reasoning phenomena (e.g., that arguments based on identity matches between a premise and a conclusion are perfectly strong).

According to two-process accounts (Darlow & Sloman, 2010; Evans, 2008; Stanovich, 2009), both heuristic and analytic processes contribute to reasoning, with each process potentially assessing an argument as strong or weak. In effect, there is more than one scale of argument strength. Both induction and deduction could be influenced by these two processes, but in different proportions, or at different times. Induction judgments would be particularly influenced by quick heuristic processes which tap into associative information about context and similarity that do not necessarily make an argument logically valid. In contrast, deduction judgments would be more heavily influenced by slower analytic processes which encompass more deliberative, and usually more accurate, reasoning. These two-process accounts have provided an explanatory framework for many results in the reasoning literature.

Surprisingly little research has directly pitted the one-process view and the two-process view of reasoning against each other. However, a study by Rips (2001) compared how people evaluate two types of arguments under either induction or deduction conditions. One type of argument was deductively correct but causally inconsistent, such as "Jill rolls in the mud and Jill gets clean, therefore Jill rolls in the mud," and the other type was deductively incorrect but causally consistent, such as "Jill rolls in the mud, therefore Jill rolls in the mud and Jill gets dirty." For both types of arguments, subjects were instructed to make either deduction judgments (respond "yes" if the argument is deductively valid) or induction judgments (respond "yes" if the argument is "strong" or highly plausible). If induction and deduction use the same information along a common scale of argument strength, then the ordering of positive responses to the two argument types should be the same for deduction and induction judgments. Rips found however, that subjects in the deduction condition gave more positive judgments to the correct but inconsistent arguments, whereas those in the induction condition gave more positive judgments to the incorrect but consistent arguments. Rips concluded that the results were evidence against a one-process account, namely the criterion-shift account, which predicts the same ordering of arguments in both conditions, with only a potential change in response bias to distinguish them.

One question about the Rips (2001) study is whether the results reveal something distinctive about the nature of induction versus deduction, or if they reveal something about causal reasoning and the use of background knowledge. It seems plausible that causal reasoning and background knowledge would play a greater role in induction. But what about situations where causal reasoning is not involved? Would there still be inherent differences between induction and deduction?

We (Heit & Rotello, 2005) conducted an experiment modeled on Rips (2001), giving either deduction or induction instructions to two groups of subjects who otherwise saw the same set of arguments. The arguments were created by modifying materials from the Rips study, in effect, stripping out their meaning so that background knowledge would not be useful. For example, "Jill rolls in the mud" was replaced with "Jill does D". The arguments were either deductively correct or deductively incorrect. Hence the experiment allowed an assessment of the criterion-shift account without the problem of the differential use of background knowledge for the deduction and induction tasks. The data were analyzed using SDT methods adapted from work on one- versus two-process theories of memory (e.g., Rotello et al., 2004). The one-process, criterion-shift account predicts that the difference in responses to correct arguments and incorrect arguments, expressed in d' units, should be the same in the deduction and induction conditions. In contrast, a substantial change in d' from induction to deduction would make it more difficult to explain deduction and induction as being based on the same scale of argument strength and varying only in response criterion.

The instructions for the deduction condition gave a brief definition of a valid argument, "assuming the information above the line is true, this **necessarily** makes the sentence below the line true". Likewise for the induction condition, there was a definition of a strong argument, "assuming the information above the line is true, this makes the sentence below the line **plausible**".

The questions were of the following form in the deduction condition.

Jill does D and Jill does R

Jill does D

Assuming the information above the line is true, does this **necessarily** make the sentence below the line true?

Circle one: VALID or NOT VALID

In the induction condition, questions were of the following form.

Jill does D and Jill does R

Jill does D

Assuming the information above the line is true, does this make the sentence below the line **plausible**?

Circle one: STRONG or NOT STRONG

Each forced-choice judgment was followed by a confidence rating, on a 1–7 scale.

For the deduction condition, the mean proportion of positive or "valid" responses for correct arguments was .89 and the proportion for incorrect arguments was .22. For the induction condition, the mean proportion of positive or "strong" responses for correct arguments was .93 and the proportion for incorrect arguments was .55. The overall proportion of positive responses was significantly higher in the induction condition than in the deduction condition, and the overall proportion of positive responses was higher for correct arguments than for incorrect arguments. There was also a significant interaction between these two variables.

In terms of sensitivity, that is, ability to distinguish between correct and incorrect arguments, the greater difference in the deduction condition suggests a greater level of discrimination. For each subject, a d' measure was calculated. The average d' was significantly higher in the deduction condition, 1.68, than in the induction condition, .93.

The next analysis used not only choice proportions but also confidence ratings, to plot ROC curves and estimate their slopes. Here, an ROC curve plots the probability of a positive ("valid" or "strong") response to valid arguments on the y-axis and to invalid arguments on the x-axis; the points indicate varying levels of confidence, with higher-confidence positive decisions appearing to the left in the space (see Macmillan & Creelman, 2005). Figure 1 shows the *z*ROC curves (normal-normal transformations of the ROCs). The curves are approximately linear, as they should be when derived from underlying Gaussian distributions of argument strength. It should also be clear that the curve for the deduction condition is more distant from the origin than is the curve for the induction condition, supporting our conclusion that sensitivity is greater for deduction. If deduction and induction had equal sensitivity and different response criteria, then the curves for the two conditions would be co-linear. The deduction instructions also led to more conservative responding, as can be seen in the leftward and downward translation of the points in that condition. Finally, it should be noted that the slopes in Figure 1 differ; the

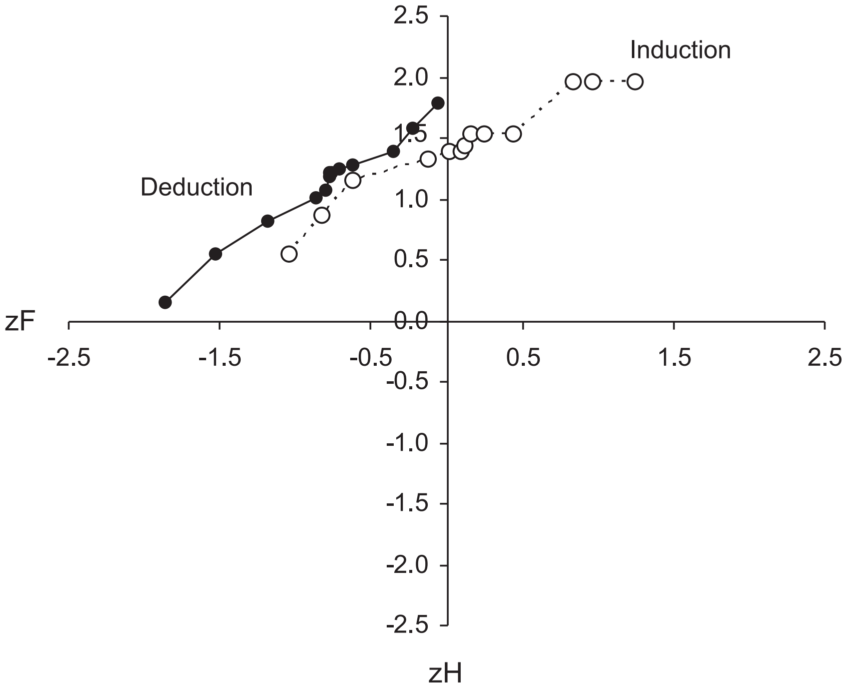

Figure 1 zROC curves from Heit and Rotello (2005), comparing deduction and induction instructions.

slope is steeper for deduction than for induction. The slope indicates the ratio of standard deviations of the invalid and valid argument distributions. This result suggests that the range of acceptable items was narrower in the deduction condition than in the induction condition.

In sum, the results were not consistent with the criterion-shift account, which would represent differences between deduction and induction solely as a change in response criterion. Instead, there were also changes in sensitivity, and in the slopes of zROC curves, that would not be predicted by the criterion-shift account. Hence, the results agree with those of Rips (2001) who also found differences between deduction and induction that could not be explained by a change in criterion.

2.1. Argument Length Effects on Inductive and Deductive Reasoning

A further parallel between approaches to assessing one- and two-process accounts of memory and reasoning involves the search for empirical dissociations between processes. An important type of evidence cited in support of two-process models of memory is that certain types of manipulations selectively affect responses based on one underlying memory process (familiarity or recollection), but have little effect on the other process. For example, reducing the time available for a decision disrupts recollection but has little effect on the accuracy of responses based on familiarity (Yonelinas, 2002).

2.1.1. Experiment

In a similar vein, Heit and Rotello (2005) found some notable differences between induction and deduction, providing evidence that these cannot be a single mechanism varying only in terms of response criterion. One key result was that the validity of an argument affects the accuracy of deductive judgments more than inductive judgments. In subsequent work we looked for further dissociations between induction and deduction; in particular we sought variables that affect induction more than deduction. One such variable is the number of premises in an argument (Rotello & Heit, 2009). Although increasing the number of premises does not itself make an argument valid, research on inductive reasoning has shown that providing more evidence can make a conclusion seem more plausible, an effect that has been referred to as "premise monotonicity" (cf., Heit, 2000; Osherson et al., 1990). Hence, we expected that increasing the number of premises would increase the perceived plausibility of invalid arguments. We also expected that increasing the number of premises would affect induction more than deduction. This prediction parallels findings in social cognition research, such that that longer communications lead to greater attitude change than shorter communications, under conditions that promote automatic or heuristic processing (e.g., Petty & Cacioppo, 1984). In contrast, we expected that deductive judgments would be more sensitive to actual validity.

Subjects saw both invalid and valid arguments, with varying argument lengths (one, three, or five premises). An example invalid argument is:

Horses have Property X
Mice have Property X
Sheep have Property X

Cows have Property X

The valid arguments were either identity matches or inclusion arguments. An example of an identity match is the following—note that one category appears both in a premise and in the conclusion.

Horses have Property X
Mice have Property X
Sheep have Property X
Rabbits have Property X
Cats have Property X

Rabbits have Property X

An example of an inclusion argument is the following—note that a premise category is a superordinate of the conclusion category.

Mammals have Property X

Horses have Property X

We assessed the proportion of positive ("strong" or "valid") responses to valid and invalid arguments (see Table 1). In general, the results pointed to two distinct effects: Validity had a greater effect on deduction judgments and argument length had a greater effect on induction judgments. Subjects distinguished between valid and invalid arguments more for deduction than for induction. As in our previous research, d' was greater for deduction (2.60) than induction (2.10). On invalid arguments, increasing the number of premises strengthened arguments significantly more in the induction condition than in the deduction condition. Interestingly, increasing the number of premises tended to weaken valid arguments, overall. Whereas invalid arguments became stronger as they got longer, valid arguments became weaker. (Liew, Hayes, and Grisham, 2012, subsequently replicated these positive and negative effects of argument length using a broader set of categories that includes artifacts.) Unlike the logical validity manipulation, argument length does not have an objective effect on the strength of an argument. Nevertheless, just as

Table 1 Proportions of Positive Responses From Rotello and Heit (2009)

Item Type	Number of Premises	Induction	Deduction
Not Valid	1	.11	.06
	3	.19	.06
	5	.28	.09
	Average	.19	.07
Valid-Identity	1	1.00	1.00
	3	.93	.97
	5	.92	.96
	Average	.95	.97
Valid-Inclusion	1	.86	.73
	3	.62	.44
	5	.66	.46
	Average	.71	.55
Valid-All	1	.96	.93
	3	.85	.84
	5	.86	.84
	Average	.89	.87

Adapted with permission

an invalid argument may seem stronger because more plausible evidence is brought to bear, valid arguments may seem more compelling, elegant, or parsimonious, and hence stronger, when they are simpler or briefer (cf., Lombrozo, 2007).

2.1.2. Modeling

These results provide a useful test bed for assessing one- and two-process models of reasoning. We first describe the two-dimensional modeling (corresponding to a two-process account). We considered the possibility that two different (orthogonal) dimensions of information were used in deduction and induction judgments. See Figure 2 for an illustration. The dimensions can be thought of as "apparent logical correctness" and "consistency with associative/background knowledge," on the principle that these would be the outputs of analytic and heuristic processing, respectively. Our starting assumptions were that valid arguments would differ more from invalid arguments along the logic axis than along the knowledge axis, and that the number of premises would influence the strength of evidence along the associative axis. Invalid arguments would be generally (but not uniformly) low on apparent logical correctness, but vary in their consistency with associative knowledge; a greater number of

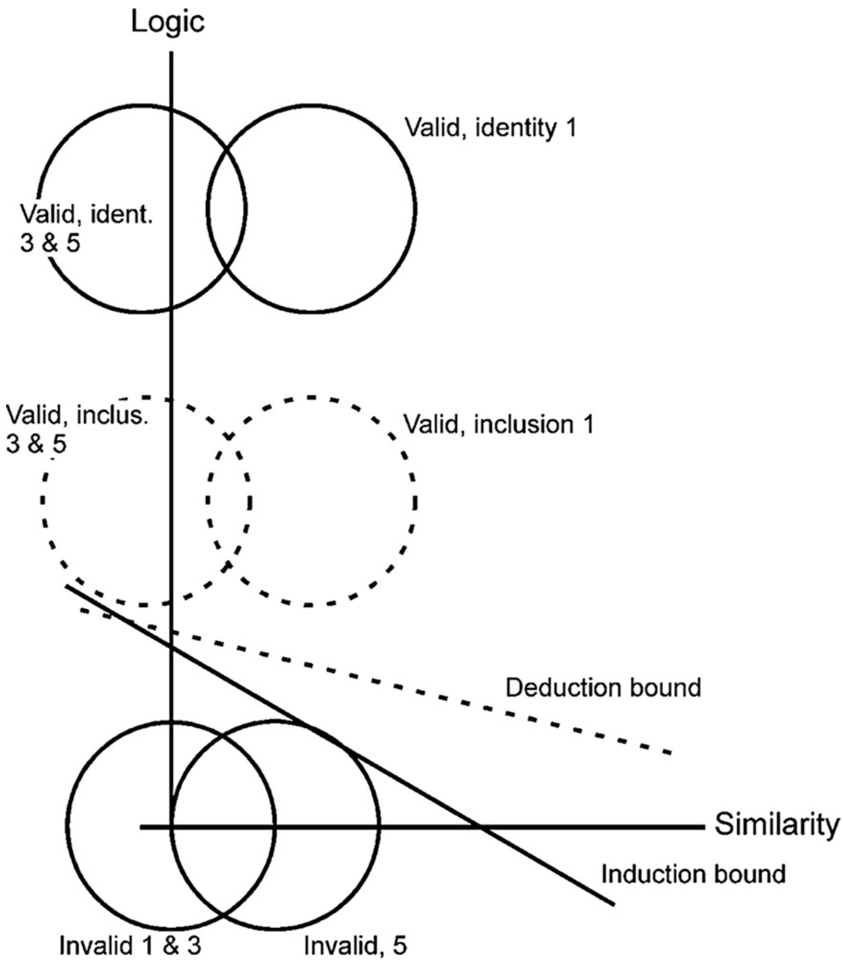

Figure 2 Two-dimensional model for both induction and deduction judgments (Rotello & Heit, 2009). (Reprinted with permission.)

premises would result in greater total similarity between premise and conclusion categories. Both valid and invalid arguments were assumed to have bivariate Gaussian distributions. To make either an induction or deduction judgment in this model, a criterion is required. We assumed that both types of judgments involve weighted combinations of evidence on the two dimensions. The relative weight given to the two dimensions is reflected in the angle of the decision bound that divides positive from negative decisions in each task: Deduction places a relatively greater emphasis on logic, and therefore the slope of that decision bound is shallower, as is shown schematically in Figure 2. Because the slope of the

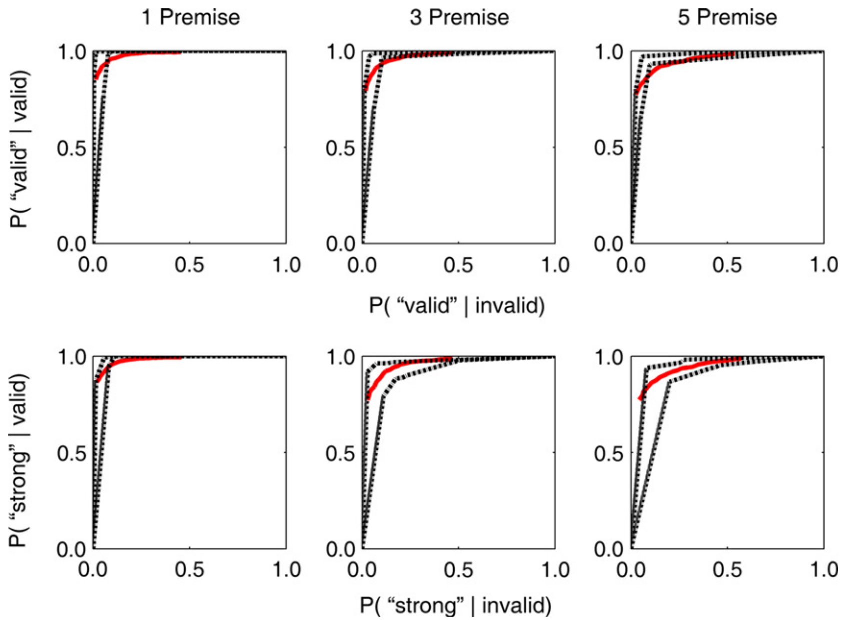

Figure 3 Simulated ROCs from the two-dimensional model (red/solid function) and 95% confidence intervals for the observed ROCs in Rotello and Heit (2009), for the identity problems. Upper row: deduction condition; lower row: induction condition. (Reprinted with permission.) (For color version of this figure, the reader is referred to the web version of this book.)

decision bound is assumed to differ for induction and deduction, and because the mean value of a distribution may be greater on one dimension than the other, this model naturally allows that accuracy for induction and deduction will differ.

We simulated this two-dimensional model over a wide range of parameter values. The predicted ROCs were created by systematically varying the intercepts of the induction and deduction decision bounds to calculate the hit and false alarm rates at a range of confidence criteria. Our goal was to demonstrate that ROCs simulated with this model would fall within the 95% confidence intervals of the observed ROCs for both induction and deduction, and for one, three, and five premise problems, assuming that the only difference between the tasks was the slope of the decision criterion. Figure 3, for identity problems, illustrates that we were reasonably successful: The dashed functions are the upper and lower bounds of the 95% CIs for each observed ROC, and the solid functions are the model-generated ROCs. In each case, the predicted ROCs generally fall within the confidence limits, and are slightly higher for the deduction than for induction, reflecting higher predicted accuracy in the deduction condition. As argument length increases, the predicted ROCs tend to shift rightward along the x-axis, more so for induction than deduction, reflecting greater tendency to respond positively to invalid arguments when they are longer. Also, as argument length increases, the predicted ROCs tend to shift downward along the y-axis, reflecting a lower tendency to respond positively to valid arguments when they are longer. The schematic two-dimensional representation in Figure 2 approximates the parameter values used to generate these ROCs.

We also tried to fit our data with a one-dimensional model like that suggested by Rips (2001): We allowed only a single scale of argument strength and assumed that induction and deduction could vary only in terms of response criterion. This model failed to capture key results. The data showed greater sensitivity to validity for deduction than induction, but the one-dimensional model predicts the same sensitivity for both. Because this model assumes that only response bias differs between induction and deduction, it predicts that the resulting data will fall on a common ROC curve on which the deduction points are each shifted to the left of the corresponding induction points, reflecting a more conservative response bias at each confidence level, but having the same sensitivity. The data also showed greater sensitivity to argument length for induction than deduction, but the model predicts the same level of sensitivity to both. In other words, this model can predict either that argument length increases the false alarm rate (as seen in the induction condition) or, if the invalid distributions are assumed to be the same regardless of argument length, that argument length does not affect the false alarm rate (as seen in the deduction condition). It cannot predict both patterns simultaneously.

2.1.3. Fluency Effects

We next set out to assess the two-dimensional model on a similar experiment, also varying validity and argument length, but with an additional manipulation. Here, fluency was manipulated by displaying the materials either in a good, readable font, or a bad, less readable font. It was expected that using a disfluent font would encourage the use of analytic processes, increasing sensitivity to validity of an argument (Alter, Oppenheimer, Epley, & Eyre, 2007). According to our two-dimensional model, an increased reliance on analytic processes would be reflected in a shallower decision slope. Only induction instructions were used, because subjects had shown such a high level of sensitivity to argument validity for deduction instructions that it seemed unlikely that sensitivity could be increased further. Our model-based prediction was that the primary difference between conditions would be in the slope of the decision bound: shallower for the bad font condition, and steeper for the good font condition. The good font was Courier New, which was the same font as used previously. The bad font was also Courier New, but was italicized and had lower contrast: The letters were a silver (gray) color rather than black.

Overall, the results were consistent with previous work, in terms of showing that making an argument longer strengthened invalid arguments and weakened valid arguments. In addition, the main prediction was supported, namely that introducing a bad font increased sensitivity to validity. We applied the same two-dimensional model to these data, varying only a few sensitivity parameters from the earlier simulations (accuracy was lower overall in this experiment). The shallower decision bound from the deduction condition of the previous experiment was associated with the bad font condition in this experiment. The model was able to capture the key differences between conditions, with no further changes needed.

2.1.4. Can People Override Argument Length Effects?

Both of these experiments showed a robust influence of argument length on inductive reasoning. We have also examined whether people can override the use of argument length (Heit & Rotello, in press). In this study, we warned subjects to not be influenced by the length of an argument. In the forewarning condition, the following additional instructions were displayed: "You will see questions that vary in length. Sometimes you will only get one sentence of information, but other times you will get several sentences of information. Note that the length of argument is irrelevant to whether it is a strong argument. Sometimes a short argument, with little information, is very convincing, and sometimes a long argument, with a lot of information, is still not very convincing at all. So in making your judgments about whether the following arguments are strong, please try to IGNORE the length of the question."

In general, because noticing argument length is automatic, because using argument length is an intrinsic part of judging plausibility, and because subjects were not given any alternative means of making the judgments, we predicted that it would be very difficult for subjects to control their own use of argument length.

The control condition replicated Rotello and Heit (2009) in terms of showing that invalid arguments are considered stronger when they are longer, and valid arguments are considered weaker when they are longer. The results of the forewarning condition were similar to the control condition. Clearly, warning subjects not to use argument length did not discourage them from doing so. It appears that argument length is such a compelling attribute, both in terms of making invalid arguments seem strong and in terms of making valid arguments seem weak, that it is very difficult to ignore. In this respect the results resemble findings from the memory literature such that exposure to lengthy lists of words promotes the false recognition of semantic associates, even when subjects are warned about this effect (Heit, Brockdorff, & Lamberts, 2004; McDermott & Roediger, 1998).

2.2. Similarity and Timing Effects on Inductive and Deductive Reasoning

In Heit and Rotello (2010), we investigated another potential dissociation, namely that similarity between premise and conclusion categories would affect induction more than deduction. We also compared speeded deduction judgments to unspeeded deduction judgments. Similarity is a central construct in some theories of inductive reasoning (Osherson et al., 1990; Sloman, 1993), and the similarity between the categories used as premises and conclusions is a key predictor of the strength of inductive judgments (see Hayes, Heit, & Swendsen, 2010, for a review). In contrast, theories of deductive reasoning typically accord little or no role to similarity, emphasizing instead the extent to which an argument is consistent with logical rules. Hence, it is valuable to compare the role of similarity in induction versus deduction. Moreover, the predicted dissociation between the effects of similarity on induction and deduction has strong parallels in the memory literature; for example, the match between the perceptual details of study and test items has a greater effect on familiarity-based responses than on recollection (e.g., Jacoby, Toth, & Yonelinas, 1993).

The analogy with memory research also suggests that manipulating decision speed may also help to identify possible multiple processes in reasoning. As noted earlier, a common finding in memory research is that familiarity-based responding is generally faster than responding based on recollection. However, there have been few such comparisons of processing speed in different types of reasoning tasks (see De Neys, 2006; Evans & Curtis-Holmes, 2005; Shafto, Coley, & Baldwin, 2007, for exceptions).

2.2.1. Experiments

In a first experiment, we manipulated both premise-conclusion similarity and the logical validity of arguments and found two dissociations: Similarity had more impact on induction, which (arguably) depended more on heuristic processing, and logical validity had more impact on deduction, which (arguably) depended more on analytic processing. The arguments were about the following kinds of mammals: bears, cats, cows, dogs, goats, horses, lions, mice, rabbits, and sheep. Arguments were assigned to a low or high similarity set based on a median split from independent assessments of similarity between premise and conclusion categories. An example (invalid) argument with high similarity is:

Horses have Property X

Cows have Property X.

Likewise, the following is an example (invalid) argument with low similarity is:

Lions have Property X

Mice have Property X.

We assessed the proportion of positive ("valid" or "strong") responses to valid and invalid arguments. As shown in Table 2, for the deduction condition, the average proportions were .94 and .04, respectively. For the induction condition, the average proportions were .95 and .12, respectively. Subjects were significantly more likely to reject invalid arguments in the deduction condition than in the induction condition, suggesting a greater influence of validity on deduction. As in prior experiments (Heit & Rotello, 2005; Rotello & Heit, 2009), d' was greater for deduction (3.31) than induction (2.82), which also suggests that deduction judgments were more affected by validity.

The data in Table 2 also suggest that inductive reasoning is more sensitive to similarity than deduction, as predicted by a two-process account. We calculated, for each subject, the difference in the positive response rate to low and high similarity invalid arguments. As expected, difference scores were significantly larger in the induction condition than in the deduction condition.

A second experiment targeted the implications of a key assumption of two-process models, namely that the analytic processes presumed to be

Table 2 Response Proportions From Heit and Rotello (2010)

Experiment	Condition	Positive Response Rate to Valid Arguments	Positive Response Rates to Invalid Arguments	
			Low Similarity	High Similarity
1	Induction	.95	.08	.14
	Deduction	.94	.04	.05
2	Fast Deduction	.83	.16	.24
	Slow Deduction	.96	.09	.11

Reprinted with permission

involved in deduction are relatively slow compared to the heuristic processes normally used in induction. Hence, if deduction judgments are speeded up, they should resemble induction judgments, because the contribution of analytic processes will be reduced. In particular, fast deduction judgments may be more influenced by similarity, and be less influenced by validity, than slow deduction judgments. In this study all subjects received deduction instructions. In the slow condition, subjects were required to wait eight seconds before making a response. In the fast condition, subjects were instructed to respond as quickly as possible and within three seconds.

We assessed the proportion of positive responses to valid and invalid arguments (see Table 2). For the slow condition, the average proportions were .96 and .10, respectively. For the fast condition, the average proportions were .83 and .20. Analogously to the deduction and induction conditions in the previous experiment, subjects were significantly more likely to reject invalid arguments in the slow condition than in the fast condition. Moreover, d' was greater for the slow condition (3.03) than the fast condition (1.80). We tested the hypothesis about the effects of similarity on deduction by calculating the difference between the positive response rates to low and high similarity invalid arguments. As predicted, responses to invalid arguments were significantly more influenced by similarity in the fast condition than the slow condition.

2.2.2. Modeling

We applied the same two-dimensional model that Rotello and Heit (2009) found to be successful, assuming that "consistency with associative/background knowledge" could reflect the similarity of the premise category to the conclusion category. For the first experiment, the sole difference between induction and deduction in the fitted model was the slope of the decision bound. Put differently, the modeling indicated that the only difference between induction and deduction is in terms of the relative weight each task assigns to information from the two dimensions. Deduction weighs the prior knowledge dimension less heavily than induction, but both types of information contribute to the judgments in each task. The simulated ROCs for the high- and low-similarity conditions are in Figure 4. Again, the model fits the data well. One key result is that when similarity is higher, the ROCs shift to the right for induction, reflecting more positive responses to invalid arguments, but much less so for deduction. In addition, the ROCs for deduction are closer to the top-left corner, reflecting greater sensitivity to validity, compared to induction. Likewise, we were able to successfully apply the two-dimensional model to the second experiment, simply by allowing the decision bound to be steeper in the fast condition, reflecting greater

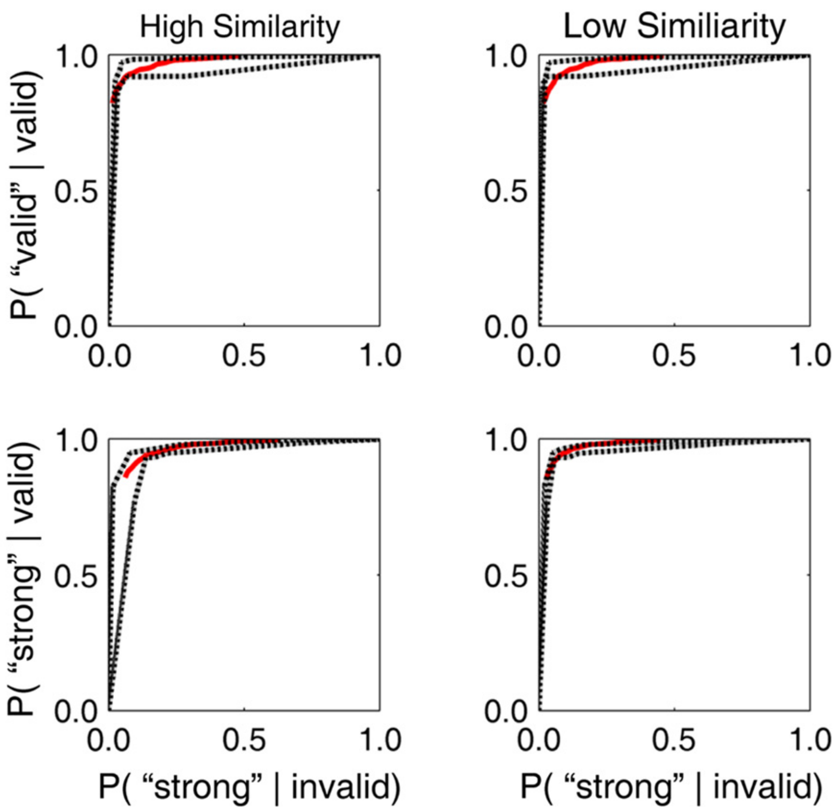

Figure 4 Simulated ROCs (solid functions) generated with the two-dimensional model, as well as 95% confidence intervals for the observed ROCs (dashed functions) for Heit and Rotello (2010). Top row: deduction condition; bottom row: induction condition. Left column: high-similarity invalid arguments; right column: low-similarity invalid arguments. (Reprinted with permission.) (For color version of this figure, the reader is referred to the web version of this book.)

weighting of the similarity information and reduced weighing of the analytic information.

We were unable to fit either experiment with a one-dimensional model in which the deduction and induction responses differed only in their response criterion, because the one-dimensional model incorrectly predicted the same effect of similarity in both conditions and because it incorrectly predicted that accuracy would not differ between conditions.

2.3. Belief Bias Effect in Deductive Reasoning

The previous two lines of work illustrated how methods that are more usual for memory research, such as SDT and ROC analyses, can be applied profitably to reasoning research. As a final example of applying memory methods to reasoning, in a study with Chad Dube (Dube, Rotello, & Heit, 2010, 2011), we addressed the important deductive reasoning phenomenon of the belief bias effect (Evans, Barston, & Pollard, 1983). Belief bias refers to the tendency to accept or reject a conclusion on the basis of its consistency with everyday knowledge, regardless of its logical status. Belief bias is typically studied using categorical syllogisms, which have a structure such as the following example.

All X are Y.
No Y are Z.

No Z are X.

The letters can be replaced with names of categories, and different configurations of quantifiers make the argument logically valid or invalid. In addition, the arguments can vary in terms of whether the conclusion is believable (e.g., All skyscrapers are buildings), or unbelievable (e.g., All buildings are skyscrapers).

Evans et al. (1983) and subsequent researchers (e.g., Evans, Handley, & Harper, 2001; Evans, Newstead, & Byrne, 1993; Markovits & Nantel, 1989; Roberts & Sykes, 2003) have reported a belief by validity interaction, such that the effect of logical validity depends on whether an argument conclusion is believable. More specifically, it has been reported that people are better able to judge whether an argument is logically valid when it is unbelievable compared to when it is believable. Hence, the belief bias effect has been measured by many researchers in terms of an interaction index, which is a difference between two difference scores: the response rate to valid unbelievable arguments minus the response rate to invalid unbelievable arguments, and the response rate to valid believable arguments minus the response rate to invalid believable arguments. This interaction index is generally reported to be positive, and is assumed to reflect greater accuracy in responding to unbelievable problems. Three decades of research on the belief bias effect, including intense theoretical development, has focused on explaining this positive interaction index (see Klauer, Musch, & Naumer, 2000, for a review).

We pointed out (Dube et al., 2010) that the interaction index parallels what in memory research would be called a recognition score "corrected" for response bias (P_r, Snodgrass & Corwin, 1988), namely the hit rate minus the false alarm rate. In effect, the interaction index is the difference between corrected scores for unbelievable arguments and for believable arguments. The corrected score assumes a particular kind of ROC, one where there is a linear relation between hits and false alarms (Swets, 1986a). Unfortunately, in memory research, ROCs tend to be curvilinear rather than linear (see Macmillan & Creelman, 2005;

Rotello, Masson, & Verde, 2008; Swets, 1986b for reviews). Using a series of simulations, Rotello et al. concluded that when ROCs are curvilinear, using corrected scores are likely to lead to faulty conclusions, such as concluding that there are differences in accuracy between conditions, when there are none. In other words, memory research using corrected scores is very much at risk.

What are the implications for reasoning research? The points made by Rotello et al. (2008) are statistical rather than being distinctly about memory, so they may also apply to the measurement of reasoning performance. Using different score measures to study the belief bias effect in reasoning is appropriate if the ROCs are linear, but is likely to lead to serious errors if the ROCs are curvilinear.

As already reviewed, we have found extensive evidence for curvilinear ROCs in both inductive and deductive reasoning tasks. In three further experiments (Dube et al., 2010), we extended this work to categorical syllogisms presented in a belief bias task. Subjects rated their confidence in more or less believable syllogisms. In a further experiment (Dube et al., 2011), we used a different procedure in which subjects made binary judgments about validity, without making confidence ratings; however they were instructed to adopt different response criteria, e.g., to respond mainly positively or negatively. In both the confidence-rating and binary cases, the ROCs were curvilinear. This finding challenges an assumption of a linear scale of argument strength, that underlies most existing work measuring the belief bias effect in reasoning.

The work of Dube et al. (2010, 2011) again shows the value of applying memory methods to reasoning. Here, we showed that many published analyses of an important phenomenon in reasoning made assumptions of linearity that were not fulfilled by the data. In addition, we developed a better way of analyzing belief bias experiments, using signal detection theory and ROCs. Our work suggests that future theoretical development should address the shift in response bias due to beliefs, rather than focusing on accuracy differences which have turned out to be a statistical artifact due to incorrect analyses.

2.4. Summary

All of the results in this section support the **analogy** view of relations between memory and reasoning. We have profited from this analogy in several different ways, drawing on memory research to illuminate reasoning. The general theoretical question, of comparing one- and two-process models, has been debated more vigorously in memory research, and we have borrowed from the extensive work in memory research that has implemented one- and two-process models and compared them directly using ROC and SDT techniques. Indeed, the very notion of treating reasoning as a signal detection task is relatively uncommon in reasoning research, and has its inspiration (for us) in memory research.

Furthermore, the manipulations that we have explained in our studies of reasoning are also classic manipulations in memory research. For example, our studies of the effect of argument length on reasoning are analogous to studies of the effect of list length on memory (e.g., Gillund & Shiffrin, 1984)—if one treats the premises in an argument as analogous to the study list in a memory experiment, and the conclusion in an argument as analogous to a test item. Research on effects of fluency on reasoning was preceded by research on fluency and memory (e.g., Whittlesea, Jacoby, & Girard, 1990). Our study of whether people can override argument length effects was directly modeled on a study of whether people can override a memory illusion (Heit et al., 2004). Although looking at similarity effects is intrinsic to studying inductive generalization, there are many corresponding studies of memory looking at similarity effects (e.g., Jones & Heit, 1993) and indeed the timing of similarity effects (e.g., Lamberts, Brockdorff, & Heit, 2003). Finally, studies of the effects of prior beliefs on reasoning are closely connected to studies of the effects of prior beliefs on memory (e.g., Bartlett, 1932; see Heit, 1997a, for a review).

At the very least, we would argue that thinking about analogies between memory and reasoning has heuristic value, in terms of suggesting theoretical and methodological approaches from one topic that can be used profitably to study the other.

3. Studying Memory and Reasoning Together

Whereas the studies described so far simply draw an **analogy** between memory and reasoning, it is possible that these two cognitive activities are even more closely related, as in the **interdependent, bigger whole, and common mechanisms** views. In this section we review our empirical work and modeling that directly examines the relations between inductive reasoning and recognition memory.

Even a cursory consideration of inductive reasoning suggests that memory should play a central role. Being able to remember the similarities (and differences) between lions and otters seems central to explaining how a property shared by these categories will be generalized. Being able to retrieve memories of dogs that are similar to the one that lives next door is useful when making predictions about that dog's behavior. Moreover, working memory seem likely to play a role when we assess the similarity between the features of things we are generalizing from and the features of things we are generalizing to (e.g., Oberauer et al, 2008).

Despite the strong case for a central role of memory in induction there have been few attempts to examine the specific connections between the processes involved in each task. Instead, memory and reasoning have been most often studied with their own experimental paradigms, addressing different questions and resulting in reasoning and memory phenomena being addressed by separate theories.

At a general level, both recognition and induction involve the generalization of knowledge about familiar stimuli to novel cases (Shepard, 1987). In the case of recognition, generalization involves a question of identity (i.e., does this novel item seem similar enough to a familiar item to conclude that it is the same thing?). Induction involves a broader form of generalization, i.e., do two items seem sufficiently similar to conclude that they will share certain novel properties? One of the key questions we address is whether there are important differences in the processes that drive induction and recognition, beyond this difference in the breadth of generalization.

A more specific point of overlap between recognition and induction is the central role accorded in each task to an assessment of the similarity between familiar and novel exemplars. In recognition, the probability that an item is recognized as "old" is a positive function of its similarity to previously studied items (Jones & Heit, 1993). Likewise induction studies have shown that the probability that a novel item is judged to have a property depends, in part, on its similarity to known instances that have that property (Hayes et al., 2010).

The proposed overlap between memory and reasoning goes beyond the level of task description; existing models of both memory (Hintzman, 1988; Ratcliff, 1990) and reasoning (Osherson et al. 1990; Sloman, 1993; see also Heit, 1997b) view similarity computation as a core process that determines performance. Despite these apparent overlaps in core processes, most models of recognition memory have not addressed reasoning and likewise previous models of inductive reasoning have not addressed memory (Heit & Hayes, 2005).

3.1. Predicting Reasoning from Memory

To explore the relationship between reasoning and memory, we (Heit & Hayes, 2011) developed a new experimental paradigm that makes reasoning and memory tasks as comparable as possible. In these experiments, subjects were either asked to make recognition judgments about a set of pictures they had studied, or make property inferences about the same set. We examined whether the overgeneralization errors that people make in visual recognition predict the pattern of generalization that other people show in inductive reasoning.

We made two general predictions. First, as noted above, we expected that reasoning and memory instructions would lead to a systematic difference between the tasks in the breadth of generalization to novel test instances. Memory instructions would emphasize that a positive response to a test item should only be made if that item has been studied. Reasoning instructions, on the other hand, would explicitly invite the subject to go beyond the information provided in inductive premises and project a property to novel items. These instructions should lead to a higher rate of positive responding to novel items under reasoning than memory conditions.

Second, if both reasoning and memory share a common underlying process, there should be a close correspondence between them in the pattern of positive responses for individual test items. Because exemplar similarity is expected to influence responses on both tasks, items that are more likely to be identified as old should generally be judged as strong candidates for property inference.

A further aim of these experiments was to examine the robustness of the relationship between memory and reasoning across a variety of task manipulations. Conventional approaches to visual recognition and induction assume that each is driven by different kinds of processes. Recognition, for example, is assumed to be strongly influenced by perceptual similarity between study and test items (Lamberts, 2002), whereas induction may involve more complex semantic or causal relations (Kemp & Tenenbaum, 2009; Medin, Coley, Storms, & Hayes, 2003; Rehder & Burnett, 2005). According to such approaches, a range of task factors might selectively affect performance on one kind of task without affecting the other. According to our approach, however, any factor that affects the specific similarity between studied items and novel test cases will affect both memory and reasoning responses in similar ways.

3.1.1. Experiment

Starting with the first experiment of Heit and Hayes (2011), we attempted to make the reasoning and memory tasks as comparable as possible. In brief, in the reasoning condition subjects were asked to learn about instances from a single category (large dogs) that shared a novel property (e.g., "has beta cells inside") whereas those in the memory condition were asked to memorize the same instances. Both groups were then shown a common test set which contained both old instances and a range of new instances that varied in similarity to old instances (i.e., unstudied dogs of various sizes). In the memory condition, people responded "yes" if they thought a test item had been presented during the study phase. In the induction condition subjects responded "yes" if they thought a test item had the target property.

The stimuli were color photographs of dogs. The same stimulus set was used for both conditions. The study list consisted of 10 pictures of large dogs, presented one at a time for 2 s each. The test list consisted of 45 pictures of dogs. There were 10 old items (the large dogs originally studied), 15 lure items (other large dogs, not previously studied), and 20 additional, new items (10 small dogs and 10 medium dogs).

The probability of responding positively to test items under memory and reasoning conditions is shown in Table 3. In the memory condition, recognition performance was good, with a relatively high hit rate on old items (studied large dogs) and a false alarm rate of .15 on new items. The false alarm rate was slightly higher on pictures of medium dogs than pictures of small dogs. For the lure items (large dogs not studied), the false alarm rate was .30.

Compared to the memory condition, subjects in the reasoning condition were more likely to give positive responses. On new items, they inferred that the dog had beta cells .45 of the time. As in the memory condition, there were more positive responses to medium dogs than to small dogs. For the lure items, the rate of positive responding was high, .68. Compared to memory, in the reasoning condition there was a higher rate of generalization, with subjects particularly likely to extend the property to the lure items that were large like the studied dogs.

To further examine patterns of generalization in memory and reasoning, a d' measure of sensitivity was calculated for each subject using individual hit rates, and false alarm rates for new (small and medium) dogs and lure items respectively. The mean sensitivity values are shown in Table 3. Sensitivity in the discrimination between old and new items was slightly higher for memory than for reasoning. Sensitivity in the discrimination between old and lure items was significantly higher for memory than for reasoning.

Next, we looked more directly at the relation between memory and reasoning. The proportion of positive responses for each of the 45 test items was averaged across subjects within each of the two experimental conditions, and the correlation between responses in different conditions was

Table 3 Results (Proportion of Positive Responses and d′) and Model Predictions, From Heit and Hayes (2011)

	Old	New Small	New Medium	All New	Lure	d' (Old–New)	d' (Old–Lure)
Empirical Results							
Memory	.68	.13	.17	.15	.30	1.50	.98
Reasoning	.82	.41	.49	.45	.68	1.03	.44
Model Predictions							
Memory	.68	.09	.18	.14	.30	1.56	.98
Reasoning	.82	.40	.50	.45	.68	1.03	.43

Reprinted with permission

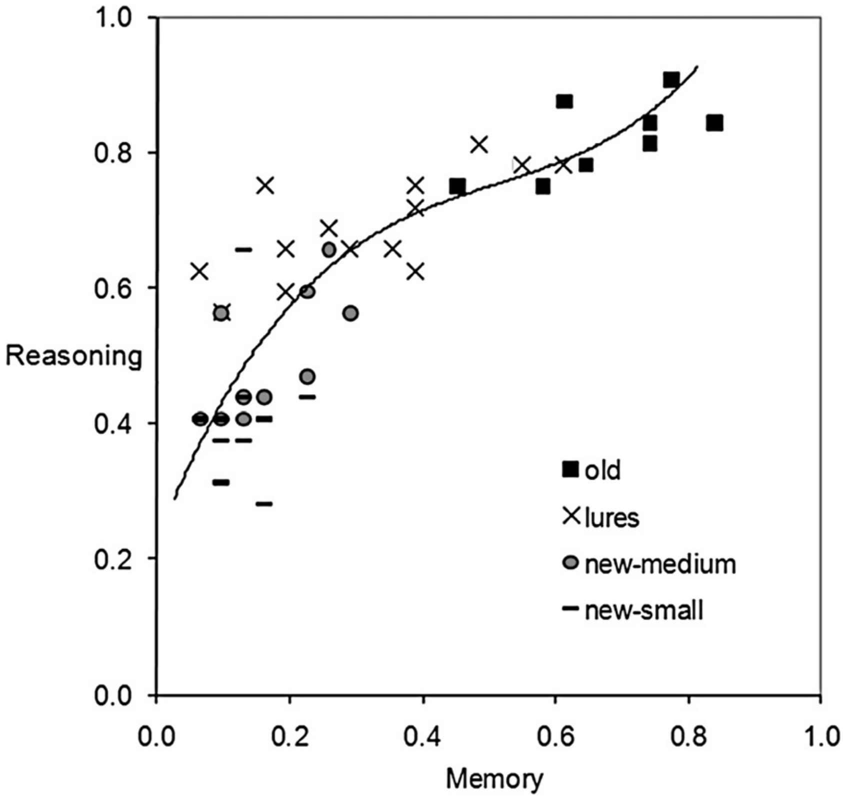

Figure 5 Scatterplots for Heit and Hayes (2011), showing memory and reasoning responses across stimuli, with trend line of GEN-EX s+d model predictions. (Reprinted with permission.)

computed. This correlation was .83. In other words, memory was a very good predictor of reasoning. This relation is illustrated in Figure 5, which shows memory responses as a function of the reasoning judgments, for both studied and new items. Note that there was a greater level of generalization for reasoning than for memory.

3.1.2. Additional experiments

Having found such a strong relation between memory and reasoning in the first experiment, in subsequent experiments (Heit & Hayes, 2011), we tried to weaken the relation by looking for variables that might affect memory and reasoning differently. In one experiment, we increased study time from 2 s per item to 5 s, on the assumption that distinctive reasoning processes might take longer to be apparent. However, the results were very similar to the first experiment.

We also tried increasing the presentation frequency of some items, because increased study frequency increases memory performance. Specifically, increasing the frequency of presentation of individual study items increases the sensitivity of old–new discrimination in recognition (e.g., Hintzman, 2001). In contrast, there is some suggestion from previous work that increasing the presentation frequency of items that share a given property might increase generalization in induction tasks (as in the premise monotonicity effects studied by Rotello & Heit, 2009). In fact, we found that frequency manipulations have similar effects on memory and reasoning, in both cases increasing the frequency of presentation of studied items increased the probability of responding positively to old items as well as the ability to distinguish old items from new items.

In another attempt to find differences between memory and reasoning, we presented, along with dogs, members of other basic level categories, namely birds and fish. Possibly, this manipulation would accentuate categories and affect category-based inductive reasoning differently than memory. However, we found that this manipulation affected memory and reasoning in a similar way: When members of other categories were presented, the key finding was that sensitivity to which items were or were not presented decreased.

We have used the same experimental paradigm in studies with five year old children (Hayes, Fritz, & Heit, 2012), observing the key results that we also found with adults, namely a higher rate of generalization for reasoning than for memory, and a positive correlation between reasoning and memory responses, albeit somewhat weaker than for adults. The main difference found between children and adults appeared to be quantitative rather than qualitative, with children showing broader generalization (or, alternatively, lower sensitivity) for both reasoning and memory.

Finally, in other work comparing memory and reasoning (Hayes & Heit, 2009), we manipulated the perceptual context in which study and test items were presented. During the study phase large dogs were presented on a background with a distinctive color and frame surrounding each picture. Test items were then presented with either context unchanged (i.e., old items were presented in their original context and new items were presented in a novel context) or context reversed (old items were presented in a novel context and new items were presented in the context that was originally used for study items). Such changes in item context generally reduce the sensitivity of old–new discrimination in memory (Murnane and Phelps, 1993; Smith & Vela, 2001). Our novel finding was that the context manipulation had a parallel effect on reasoning; people were more likely to generalize a property to test items (and showed reduced old–new sensitivity) when there was a change in item context between study and test.

Despite great efforts, we have struggled to find major differences between memory and reasoning, other than reasoning having a broader generalization gradient than memory. Across the eight experiments in Heit and Hayes (2011), the average correlation between memory and reasoning responses was .87. Somewhat to our surprise, when examining memory and reasoning responses to a common set of stimuli, we found a great deal of similarity and few differences.

3.2. An Exemplar Model of Memory and Reasoning

Given the great deal of similarity we observed between memory and reasoning, we set out to develop a model of both cognitive activities. An important goal of this work was to examine whether reasoning and memory performance could be accommodated within a single computational model. The core assumption of our model of reasoning and memory is adapted from the generalized context (GCM) family of categorization models (e.g., Nosofsky, 1986, 1988). We assume that the tendency to make a positive response to a test stimulus in memory and reasoning is a positive function of the total similarity between that stimulus and all studied items. Exemplar models have been successful in accounting for patterns of categorization and recognition of the same stimulus sets (e.g., Shin & Nosofsky, 1992) but have only rarely been applied to inductive reasoning data (e.g., Estes, 1994; Heit, 1992). An important advantage of exemplar models is that they can account for empirical dissociations between tasks without assuming multiple cognitive systems (e.g., Nosofsky & Zaki, 1998).

In our own work comparing reasoning and memory, we allowed for the possibility of different response criteria, but we also investigated the possibility that the shape of the generalization gradient will be different for reasoning and memory. The key prediction is that reasoning would involve broader generalization to novel items whereas memory would be more sensitive to exact matches between studied and test items. After all, the nature of the task of inductive reasoning is to generalize to unseen instances (e.g., Heit, 2007), whereas the goal of a recognition memory task is to reject unseen instances.

The basic version of GEN-EX (so named because it GENeralizes from EXamples) is embodied by two equations. Eq. (1) shows the familiarity rule: the familiarity of each test stimulus equals its summed similarity to the n studied items. Similarity is assumed to be a negative exponential function of distance between the test and study items, calculated according to the standard Euclidean formula. The free parameter c reflects specificity of responding to test items; lower values of c

correspond to broader generalization while higher values correspond to narrower generalization gradients.

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

$$\mathrm{resp}(\mathrm{test}) = \frac{\mathrm{fam}(\mathrm{test})}{\mathrm{fam}(\mathrm{test}) + \beta} \tag{2}$$

The response rule is shown in Eq. (2). 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.

The central predictions of Heit and Hayes (2011) were that positive responses under recognition and induction instructions should be well predicted by GEN-EX using the same old–new similarity values for test items but that the c parameter should reflect broader levels of generalization under induction as compared with recognition conditions.

To model recognition and induction judgments using GEN-EX, we collected similarity ratings between pairs of study and test items, from another group of subjects. There were 10 study items, and 45 test items, giving a total of 450 pairs. It was assumed that similarity would be a negative exponential function of psychological distance (e.g., Nosofsky, 1986, 1988), as illustrated by Eq. (3).

$$\mathrm{sim}(x, y) = \exp(-c \ \mathrm{dist}(x, y)) \tag{3}$$

There were three c parameters. One was used for converting similarity ratings to distances, as in Eq. (3). That is, Eq. (3) was solved for distance as a logarithmic function of similarity, to calculate distances as a function of an estimated c parameter. The other two were used for converting distances back to similarity, as in Eq. (1), when calculating familiarity for recognition and for induction. In addition, there were two β parameters, serving as scaling parameters for recognition and induction.

We developed multiple versions of GEN-EX, to take account of other possible response patterns not predicted by item similarity. Loosely speaking, these patterns could be thought of as deterministic or rule-based responding. One pattern would be to form a sub-category corresponding to the studied items. Here, the sub-category would be large dogs. A subject responding according to sub-categories would respond positively to large dogs, whether old or lures, and respond negatively to small and medium dogs (see Hayes, Foster, & Gadd, 2003, for related ideas).

Another possible pattern would be to respond deterministically depending on whether the item was actually studied, as if memory were perfect.

A subject responding deterministically would respond positively to old items and negatively to new items, and would not be willing to generalize. In terms of memory models, this pattern is equivalent to recollection (Yonelinas, 2002). GEN-EX d was developed to examine this kind of pattern.

The most complete version of the model, GEN-EX s+d, had free parameters corresponding to both of these response patterns as well as responding according to item similarity, as show in Eq. (4).

$$\mathrm{resp}(\mathrm{test}) = s \cdot \mathrm{large}(\mathrm{test}) + d \cdot \mathrm{old}(\mathrm{test}) + (1 - s - d)\frac{\mathrm{fam}(\mathrm{test})}{\mathrm{fam}(\mathrm{test}) + \beta} \quad (4)$$

Here, s represents the probability of responding according to sub-categories (responding positively whether the test item is a large dog). Likewise, d represents the probability of responding deterministically according to whether the item is old. Distinct s and d parameters were estimated for recognition and induction, to allow for different response patterns for the two tasks. For example, it seemed possible that induction might entail more sub-category based responding, and recognition would entail more deterministic (i.e., recollective) responding. There were also three restricted models. In GEN-EX s, the d parameter was set to zero, so there was no deterministic responding. In GEN-EX d, the s parameter was set to zero, so there was no sub-category responding. Finally, in GEN-EX null, both the s and d parameters were set to zero; this is the original GEN-EX model.

The four GEN-EX models were used to generate predictions for 90 data points, corresponding to the 45 test items for memory and for inductive reasoning respectively. Although all four versions of the GEN-EX model fit the data reasonably well, the GEN-EX s+d model fit significantly better than any of the restricted models, after taking account of number of parameters (having a correlation of .94 with the data across 90 data points, using just 9 free parameters). Hence, both the sub-category responding and deterministic responding components led to significant improvements in the fit of the model.

Table 4 shows the estimated parameter values for GEN-EX s+d. Note that, as predicted, the c value was considerably higher for memory than for reasoning, reflecting narrower generalization for memory and broader generalization for reasoning. The s and d parameters were estimated to be non-zero but fairly low, so that the greatest overall influence on responses was item similarity rather than subtyping or deterministic recollection. The value of s was estimated to be higher for induction than for recognition, suggesting a greater influence of sub-categories for induction. In contrast, d was estimated to be higher for recognition than for induction, suggesting a greater tendency to respond deterministically for recognition. Finally, the

Table 4 Summary of Model Fitting for GEN-EX s+d, From Heit and Hayes (2011)

Data Points	90
RMSE	.0859
Correlation	.9409
χ^2 over s model	7.45*
χ^2 over d model	6.86*
χ^2 over null model	12.35*
c-sim	2.18
c-rec	3.56
c-ind	.91
β-rec	.79
β-ind	.66
s-rec	.01
s-ind	.11
d-rec	.17
d-ind	.09

Note: The table shows goodness of fit (root mean squared error and correlation) for the GEN-EX s+d model, comparative tests showing improvement in goodness of fit over the s, d, and null models, and estimated parameter values for the s+d model. See text for definitions of parameters.*$p < .05$.
Adapted with permission

β parameter values are similar for memory and reasoning; there is little evidence for different response scaling between tasks.

Table 3 shows average predictions of the GEN-EX s+d model, for key types of stimuli. The table shows that the main trends in the data have been captured, such as differences between memory and reasoning conditions, and differences between old, lure, new medium, and new small items. Likewise the predicted d' measures are close to the original results. Note that these are only derived measures; the simulation made separate predictions for each of the 45 test items in both the memory and reasoning conditions. In the simulation, the average predicted correlation between memory and reasoning was .92. In Figure 5, model predictions are shown as a trend line, which well captures the pattern in the data points. One key trend is that the model predicts a higher rate of positive responding for the reasoning condition than for the memory condition, even for new and lure items, a consequence of the lower c parameter for reasoning, leading to higher familiarity values in Eq. (1).

We also fit the GEN-EX model to the other seven experiments in Heit and Hayes (2011). In general, the basic GEN-EX model, without deterministic or subtyping components, gave a good account, simply by assuming a broader generalization gradient for reasoning than for memory. Adding the deterministic, recollection-like component tended to improve the model fit somewhat, but there was little evidence overall of the need for a subtyping component.

3.3. ROC Analyses of Memory and Reasoning

In Section 2 of this chapter, we noted that differences in sensitivity between induction and deduction suggest that these two forms of reasoning cannot depend on a single underlying process. A sharp-eyed reader may have noticed that in Section 3, we have explained differences in sensitivity between memory and reasoning mainly in terms of a single process, exemplar similarity, varying in terms of the generalization gradient. So is there a contradiction between Section 2 and Section 3? We note that we did not simply find sensitivity differences between induction and deduction; we also found other results pointing to two underlying processes, such as dissociations based on length and similarity, and other effects of fluency and timing. In contrast, we found extensive similarities, and no dissociations, when comparing memory and inductive reasoning. Indeed, the best fitting model for the experiments in Section 3 included not only a similarity-based process, but also rule-based processes, for both memory and reasoning. So the evidence is considerably different for induction versus deduction on the one hand, and memory versus reasoning, on the other.

Still, it is somewhat counterintuitive that an exemplar model can explain sensitivity differences based on just a single cognitive process. It is interesting to ask whether exemplar modeling and ROC analysis would lead to different conclusions when both are applied to a common data set. We now present a previously unpublished experiment, that brings together the two strands of this chapter, namely applying memory methods such as ROC analysis to reasoning (Dube et al., 2010, 2011; Heit & Rotello, 2005, 2010; Rotello & Heit, 2009), and simultaneously studying, and modeling, reasoning and memory (Heit & Hayes, 2011). In the Heit and Hayes experiments, we did not collect confidence ratings, so it was not possible to perform ROC analysis. However, in this study, we did so.

This experiment was same as the first experiment of Heit and Hayes (2011), in terms of the design, stimuli, and procedure, except that subjects made confidence ratings on a 1–3 scale (1 = 'sure'; 3 = 'guess') after each judgment, and there were four main conditions rather than two. In addition to varying whether subjects performed a reasoning task or a memory task, we also varied whether subjects saw the original pictures of the dogs or if they saw "noisy" versions of the dogs, with lowered contrast and added perceptual noise. The noise manipulation was essentially another attempt to find a variable that might affect reasoning and memory differently.

The control conditions replicated previous results from Heit and Hayes (2011). We again found that subjects were more sensitive to the difference between old and new items for memory ($d' = 1.25$) than for reasoning ($d' = 0.94$). Although the noisy dog conditions were conceptually analogous to the bad font condition of Rotello and Heit (2009), here degrading the stimuli made performance worse. In the noisy conditions, d' for memory was 1.02 for memory and .48 for reasoning. Hence, there were main effects of going from memory to reasoning, and of adding noise, with both manipulations reducing sensitivity. There were strong correlations between memory and reasoning judgments in both the control ($r = .92$) and noisy ($r = .88$) conditions.

We applied the GEN-EX model to these data. Interestingly, this dataset did not show significant evidence for the deterministic or subtyping components of the model (recall that evidence for these processes was mixed across the eight experiments in Heit and Hayes, 2011). However, the original GEN-EX model gave a satisfactory fit to the data, e.g., the correlation across 180 items model predictions and data was .84, with 10 free parameters. The key estimated parameters were the c values, corresponding to steepness of the generalization gradient. These were 7.92 for memory-control, 5.78 for reasoning-control, 6.45 for memory-noisy, and 3.00 for reasoning-noisy. The GEN-EX model was able to account for differences between control and noisy dogs, as well as between memory and reasoning, simply by lowering the c parameter, that is, by broadening the generalization gradient. Note that there were small differences in the β parameter across conditions, but these did not seem to be the source of the greater level of positive responding for reasoning compared to memory.

Next, we examined the full dataset, including the confidence ratings, with the goals of assessing accuracy and response bias differences across conditions. In addition, because ROCs have been used in the memory literature to assess the contributions of a threshold recollection process to recognition (e.g., Yonelinas, 1994), the full data provide another window onto the potential contributions of recollection to reasoning.

The first analysis focused on A_z, which measures the area under the ROC; it takes on the value 0.5 for chance performance and 1 for perfect performance. The results, shown in Table 5, are consistent with what is visually clear in Figure 6: memory accuracy is greater than reasoning accuracy, and perceptual noise decreases accuracy in both tasks.

Figure 6 also shows that the operating points on the reasoning ROCs are shifted to the right along their respective curves, compared to the memory ROCs. These shifts suggest potential response bias differences, with reasoning leading to more liberal responding. To assess these differences, we computed the response-bias measure β, which is the likelihood ratio of the target and lure distribution ordinates at the

Table 5 Sensitivity and Bias For Memory and Reasoning Instructions, Using Control and Noisy Dogs Stimuli

		A_z	β
Memory	Control	.79	1.26
	Noisy	.76	1.12
Reasoning	Control	.75	.97
	Noisy	.64	.90

decision bound (this value is not the same as the β in Eq. (2) of the GEN-EX model, although it plays a related role in determining response tendencies). β equals 1 at the intersection of the distributions, increases in value for conservative criteria, and decreases for liberal criteria. β is particularly useful in experiments such as ours, in which the number of

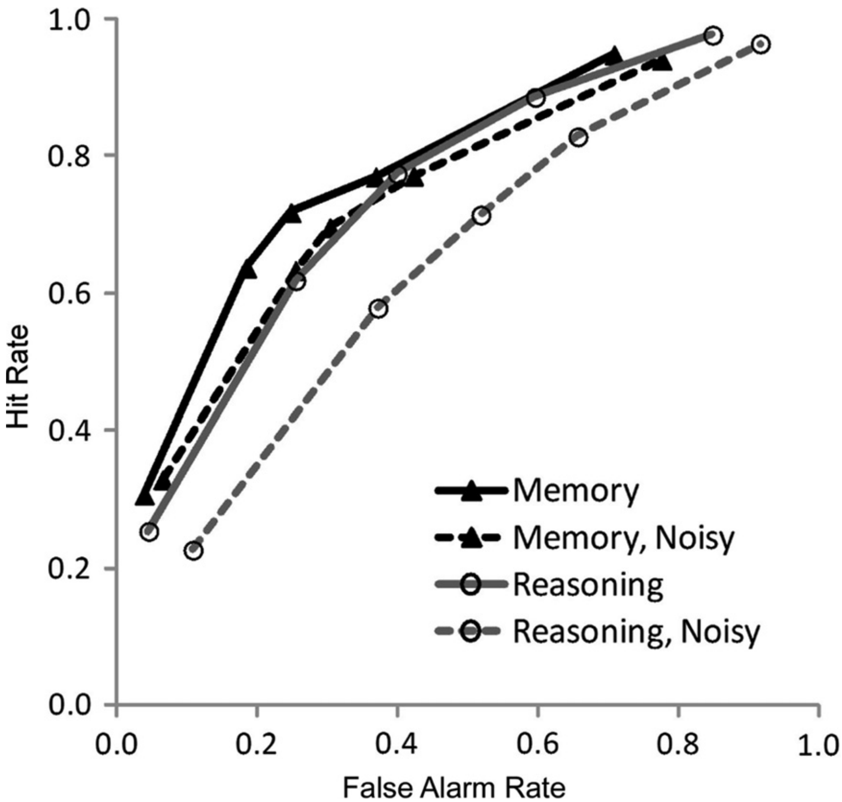

Figure 6 Empirical ROC curves for memory and reasoning instructions using control and noisy dog stimuli.

studied items and lures are not equated on the test list. In contrast to the GEN-EX analyses, which found no systematic effect of condition on the response bias parameter, the ROC based analysis pointed to a difference in response bias parameters for memory versus reasoning; subjects were significantly more liberal in the reasoning task. It appears that in the GEN-EX model, changes in the c parameter may be able to account for not only sensitivity changes but also some bias changes. For the ROC analysis, neither perceptual noise, nor the interaction of noise with task, had significant effects on β.

Our next analyses addressed whether the ROCs revealed any evidence for the contribution of a recollection process to either memory or reasoning judgments. We considered both high-threshold and continuous versions of recollection. In the threshold version, only studied items can be recollected; new items cannot. Yonelinas (1994) proposed this highly influential two-process model for recognition memory as a way of accounting for the asymmetry in memory ROCs. Specifically, this model predicts that the hit rate (H) and false-alarm rate (F) vary as a function of response criterion, c, according to Eqs (5) and (6):

$$H = R + (1 - R)\Phi(d' - c) \tag{5}$$

$$F = \Phi(-c) \tag{6}$$

where R is the probability of recollection and Φ is the cumulative distribution function for a normal distribution. We applied this model to all four ROCs simultaneously, assuming that all parameters were free to vary across conditions, or that some were shared across conditions. The best-fitting model was chosen to minimize the fit statistic G^2, which has a χ^2 distribution with degrees of freedom equal to the number of free parameters. Differences in G^2 across various parameter constraint options also follow a χ^2 distribution with degrees of freedom equal to the difference in number of free parameters. The best fit overall was provided by the model that assumed $R = 0$ in all conditions ($G^2_{\mathrm{df}=6} = 24.08$, $p = .09$), which didn't differ significantly from the model in which all parameters were free to vary ($G^2_{\mathrm{df}=12} = 20.31$, $p = .06$; $\Delta G^2_{\mathrm{df}=4} = 3.77$, $p = .44$). Constraining the familiarity parameters to be equal across tasks ($F_{\mathrm{reas}} = F_{\mathrm{mem}}$; $F_{\mathrm{reas,noise}} = F_{\mathrm{mem,noise}}$) significantly reduced the fit ($\Delta G^2_{\mathrm{df}=2} = 9.98$, $p < .01$), as did constraining the familiarity parameters to be equal across the noise variable ($\Delta G^2_{\mathrm{df}=2} = 19.49$, $p < .001$). Thus, the ROC based analyses are consistent with the results of the GEN-EX model in concluding that recollection did not make a substantial contribution to either the memory or reasoning judgments.

We also applied a continuous recollection model to the memory and induction data. Several authors have identified problems with threshold models (e.g., Kinchla, 1994; Dube et al., 2010, 2011; Dube & Rotello, 2012; Pazzaglia, Dube, & Rotello, 2012) as well as of the threshold component of the two-process model (e.g., Dougal & Rotello, 2007). Recently, Wixted and Mickes (2010) have argued that recollection is a continuous process rather than a discrete, threshold process. Their continuous dual-process model (CDP) assumes that both recollection and familiarity can contribute to every memory judgment: The two memory signals are summed, yielding an evidence distribution that resembles the predictions of a one-dimensional unequal-variance signal detection model. To fit the CDP model, one needs estimates of the contribution of recollection and familiarity processes, which could come from remember-know judgments or from source memory decisions. Lacking either of those additional forms of data, any attempt to fit the CDP to our data would simply amount to fitting an unequal-variance model. For this reason, we adopted a different approach, which was to replace the dual-process model's threshold recollection process with a continuous recollection process based on signal detection theory. The basic equations for the dual-process model (Eqs (5) and (6)) were modified to assume that recollection is continuous (hence, the RC subscript),

$$R = \Phi\left(d'_{\mathrm{RC}} - C_{\mathrm{RC}}\right) \tag{8}$$

to allow lures to be (falsely) recollected,

$$F = R + (1 - R)\Phi(-C), \tag{9}$$

and to allow for the possibility that the underlying distributions were unequal in variance. (The latter change allows the model to fit asymmetrical ROCs.) This model is similar to the process-pure model for remember-know judgments discussed by Macmillan and Rotello (2006).

This continuous recollection model was fit to the data using the same methods as before. Once again, the best-fitting variant of the model assumed that there was no recollection involved in either memory or reasoning judgments ($G^2_{\mathrm{df}=17} = 17.03$, $p = .20$); this variant did not fit worse than the version in which all parameters were free to vary across conditions ($G^2_{\mathrm{df}=13} = 16.60$, $p = .22$; $\Delta G^2_{\mathrm{df}=2} = .43$, $p = .98$). In terms of the Akaike (1973) Information Criterion (AIC), comparing the fit of the continuous recollection model to the threshold version confirmed that they are indistinguishable (AIC = 17146 and 17143 for threshold and continuous versions). Overall, the ROC based analyses support those based on the GEN-EX model: All analyses indicate that memory accuracy was greater than reasoning accuracy, and that adding perceptual noise to the stimuli impaired performance on both tasks. No recollection or sub-typing was needed to fit the data, and the ROC analyses implied that subjects responded more liberally in the reasoning task.

3.4. Summary

The research reviewed in this section establishes strong empirical connections between memory and reasoning, e.g., strong correlations in responses over a common set of stimuli when people perform one task or the other. Furthermore, we have shown that it is possible to apply the same model to both tasks, with a quantitative change in parameter values between tasks rather than a qualitative change. We have made this point with two types of modeling, exemplar models and signal detection models. At this point, we see the models as complementary, and we would not attempt to choose one over the other. The larger point is that the empirical similarities and the potential for a common modeling framework strongly suggest that the relation between memory and reasoning is more than an **analogy**.

Although we generally favor the point that memory and reasoning are **interdependent**, this view was not addressed directly here. In Heit and Hayes (2011, Experiment 1D), we did consider the possibility that the strong correlations we had found between memory and reasoning performance was an artifact of the reasoning task requiring memory retrieval. That is, making a judgment about a test item, in the reasoning condition, still required retrieving studied items from memory. To address this issue, we ran an additional reasoning condition that had no memory component. In this case, study phase stimuli were available for inspection during test. Not surprisingly, responses to old items were close to ceiling. However, in other respects the results were comparable to the other experiments, e.g., there was a strong overall correlation between memory and reasoning responses. This results suggest that the close relation between memory and reasoning found in our studies was not an artifact of the reasoning task requiring a judgment from memory.

The close empirical connections we found between memory and reasoning are consistent with both the **bigger whole** and **common mechanisms** views, because different cognitive activities that are carried out by the same architecture, or by a common set of underlying mechanisms, could be expected to show various regularities. Although exemplar models and signal detection models may not get the same credit as Bayesian models or neural networks for being general cognitive architectures, we see the success of both exemplar models and signal detection models here as broadly consistent with the **bigger whole** view. We would expect that models based on other cognitive architectures could also be applied successfully to our results. The modeling we performed provides insights into the **common mechanisms** view as well, specifically pointing to a strong role for generalization in both memory and reasoning, as well a secondary role for something like a rule-based mechanism based on recollection. At this point, we would not favor either the **bigger whole** or **common mechanisms** over the other; each seems to provide some unique insights.

4. Conclusion

4.1. Relations to Other Cognitive Activities

We have previously argued (Heit & Hayes, 2005) that the lines between different cognitive activities are often blurry. This point applies not just to memory and reasoning, but to other cognitive activities such as categorization (which also is typically treated in a different chapter in cognitive psychology textbooks). For example, categorization would allow people to link an individual item to a reference class, such as categorizing some animal as a dog. Inductive reasoning allows people to infer further properties based on category membership, such as that a dog will bark. Recognition memory allows people to judge whether an individual has been observed before, such as whether this particular dog has been seen before. These three cognitive activities are so similar that the defining line between them is sometimes obscure. For example, the boundary between properties and categories can be blurred, so that barking can be considered a property but animals that bark can be considered a category (cf., Anderson, 1991; Billman & Heit, 1988; Heit, 1992). Hence whether a particular inference should be called reasoning or categorization is sometimes unclear. In addition, recognition memory judgments themselves can be considered as a kind of categorization judgment, in which an individual is assigned either to the category of things previously observed or to the category of novel items. We see these fuzzy boundaries between categorization, reasoning, and memory not as a problem but rather as an opportunity. That is, the close relations among these activities invite a common psychological account that addresses all three.

Our argument for close links between reasoning and memory is also motivated by previous work that has discovered strong empirical and theoretical links between inductive reasoning and categorization (Rehder & Burnett, 2005; Sloutsky & Fisher, 2004) as well as categorization and recognition memory (Estes, 1994; Nosofsky, 1988). Rehder and Burnett (2005) reported a strong empirical relationship between reasoning and categorization, with correlations across items of 0.70–0.99 between inferences about whether an exemplar possessed a novel property and category membership judgments for the same items. Sloutsky and Fisher (2004) also observed a strong correlation across items between children's category membership and judgments about the generalization of novel properties.

With regard to the relationship between categorization and recognition, exemplar models of categorization have made the case for a systematic relationship between performance on these two tasks. In support of this argument, Nosofsky (1988) as well as Nosofsky and Zaki (1998) showed that that old–new recognition judgments and categorization judgments of the same items can be explained by assuming that both rely on a common memory trace for exemplars but that different decision rules are applied in each task. What is particularly notable about this work is that it contrasts with research that asserts categorization and memory depend on different processes or even different systems (e.g., Knowlton & Squire, 1993; Smith & Minda, 2001).

4.2. What is the Relation between Memory and Reasoning?

Finally, we return to our original question from the beginning of this chapter: What is the relation between memory and reasoning? We believe that our empirical and modeling results make a strong case that there is a close **analogy** between memory and reasoning that is very useful for guiding research. When it comes to experimental methodology, analyses, and theoretical ideas, reasoning researchers have a lot to learn from memory researchers, and vice versa.

We also believe that our results point to commonalities between memory and reasoning that are much deeper than an **analogy**, and are well explained in terms of a memory and reasoning being part of a **bigger whole** or sharing **common mechanisms**. It would be intriguing to compare memory and reasoning tasks using functional neuroimaging, taking advantage of the logic of forward inference (Henson, 2006), to seek additional evidence bearing on this issue. For example, if there are qualitative differences in brain activity for memory and reasoning tasks while holding other variables constant (materials, timing, task difficulty, etc.) that would suggest some role for distinct mechanisms underlying each task. To our knowledge such a study has not been done, but see Nosofsky, Little, and James (2012) for a related experiment comparing memory and categorization.

The disparate results presented in this chapter can be explained in terms of two underlying mechanisms: a faster process that depends on generalization, familiarity, and associations, and a slower process that depends on recollection and rule-following. The first of these processes was found to be a core component of memory and reasoning judgments in every study we have run. Some evidence for the second, more deliberative component was also found, although it seems to play a relatively minor role in induction. An important goal for future work is to examine when the slower deliberative processing component plays a more prominent role in deductive reasoning and in induction involving more complex relations (cf. Medin et al., 2003; Rehder & Burnett, 2005).

We suggest that such a division between faster and slower cognitive processes has a greater reality than the traditional division between memory and reasoning (see Kahneman, 2011, for a very broad review of fast and slow thinking). In essence, we are suggesting a different way of carving up the cognitive pie. Dividing cognition into separate domains such as memory, reasoning, categorization, and so on, may have pedagogical value (e.g., when writing textbooks or teaching classes). But it is important to keep in mind that the separations between cognitive activities are largely due to pre-theoretical assumptions and socially constructed conventions (cf., Kuhn, 1996), rather than direct empirical comparisons or attempts to model underlying processes. For example, the influential list of cognitive activities by James (1890) was derived by intuition about functions rather than by experimentation. Put another way, using the labels "memory" and "reasoning" to describe distinct cognitive activities is a matter of terminology but not necessarily reality.

It may be difficult to imagine a future where psychologists, rather than attending separate conferences on either reasoning or memory, attend conferences on either fast or slow cognition. Indeed, we would argue that researchers studying fast and slow cognition, as well as generalization and recollection, should be attending the same conferences, and that in many cases these should be the same researchers. We hope that our own work encourages other researchers to investigate and indeed to reconsider the relations among various cognitive activities.

ACKNOWLEDGEMENTS

This work was supported by Australian Research Council Discovery Grant DP0663963 to Brett Hayes and Evan Heit, and by National Science Foundation (US) grant BCS-0616979 to Evan Heit and Caren Rotello. We thank Rick Dale for comments on a previous version of this chapter. This work has also benefited from comments on presentations given at Stanford University and Washington University in St. Louis.

REFERENCES

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In B. N. Petrov, and F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*, 569–576.

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Barsalou, L. W., Breazeal, C., & Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processes*, *8*, 79–81.

Bartek, B., Lewis, R. L., Vasishth, S., & Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1178–1198.

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

Billman, D., & Heit, E. (1988). Observational learning without feedback: A simulation of an adaptive method. *Cognitive Science*, *12*, 587–625.

Brainerd, C. J., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, *17*, 210–247.

Brainerd, C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100*, 42–67.

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 811–823.

Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 382–392.

De Neys, W. (2006). Automatic–heuristic and executive–analytic processing during reasoning: Chronometric and dual-task considerations. *The Quarterly Journal of Experimental Psychology*, *59*, 1070–1100.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*, 523–533.

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review*, *14*, 423–429.

Dube, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 130–151.

Dube, C., Rotello, C., & Heit, E. (2010). Assessing the belief bias effect with ROCs: It's a response bias effect. *Psychological Review*, *117*, 831–863.

Dube, C., Rotello, C., & Heit, E. (2011). The belief bias effect is aptly named. *Psychological Review*, *118*, 153–163.

Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, *115*, 426–446.

Estes, W. K. (1994). *Classification and cognition*. Oxford University Press.

Evans, J. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306.

Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, *11*, 382–389.

Evans, J. S. B. T., Handley, S. J., & Harper, C. N. J. (2001). Necessity, possibility and belief: A study of syllogistic reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *54*, 935–958.

Evans, J. S. B. T., Newstead, S. E., & Byrne, R. M. J. (1993). *Human reasoning: The psychology of deduction*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*, *58*, 19–34.

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, Mass: MIT Press.

Gardiner, J. M., Ramponi, C., & Richardson-Klavehn, A. (2002). Recognition memory and decision processes: A meta-analysis of remember, know, and guess responses. *Memory*, *10*, 83–98.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1—6.
Hayes, B. K., Foster, K., & Gadd, N. (2003). Prior knowledge and subtyping effects in children's category learning. *Cognition*, *88*, 177—199.
Hayes, B. K., & Heit, E. (2009). *Exploring the Relationship Between Inductive Reasoning and Recognition Memory*. Boston, USA: Paper presented at the 50th Annual Meeting of the Psychonomic Society.
Hayes, B. K., Fritz, K., & Heit, E. (2012). The relationship between memory and inductive reasoning: Does it develop? *Developmental Psychology*.
Hayes, B. K., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 278—292.
Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, *24*, 341—380.
Heit, E. (1997a). Knowledge and concept learning. In K. Lamberts, and D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 7—41). London: Psychology Press.
Heit, E. (1997b). *Features of similarity and category-based induction. Proceedings of SimCat 97: An interdisciplinary workshop on similarity and categorisation* (pp. 115—121). Department of Artificial Intelligence, University of Edinburgh.
Heit, E. (2000). Properties of inductive reasoning. *Psychonomic Bulletin & Review*, *7*, 569—592.
Heit, E. (2007). What is induction and why study it? In A. Feeney, and E. Heit (Eds.), *Inductive reasoning* (pp. 1—24) Cambridge: Cambridge University Press.
Heit, E., Brockdorff, N., & Lamberts, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review*, *11*, 380—386.
Heit, E., & Hayes, B. K. (2005). Relations among categorization, induction, recognition, and similarity. *Journal of Experimental Psychology: General*, *134*, 596—605.
Heit, E., & Hayes, B. K. (2008). Predicting reasoning from visual memory. In: *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 83—88).
Heit, E., & Hayes, B. K. (2011). Predicting reasoning from memory. *Journal of Experimental Psychology: General*, *140*, 76—101.
Heit, E., & Rotello, C. M. (2005). Are there two kinds of reasoning? In: *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.
Heit, E., & Rotello, C. M. (2010). Relations between inductive reasoning and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 805—812.
Heit, E., & Rotello, C. M. (in press). The pervasive effects of argument length on inductive reasoning. *Thinking & Reasoning*.
Henson, R. N. (2006). Forward inference in functional neuroimaging: dissociations vs associations. *Trends in Cognitive Science*, *10*, 64—69.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528—551.
Hintzman, D. L. (2001). Judgments of frequency and recency: How they relate to reports of subjective awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 1347—1358.
Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General*, *122*, 139—154.
Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, USA*, *105*, 6829—6833.
James, W. (1890/1983). *Principles of psychology*. Cambridge: Harvard University Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, *50*, 189–209.
Jones, C., & 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.
Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus, and Giroux.
Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.
Kinchla, R. A. (1994). Comments on Batchelder and Riefer's multinomial model for source monitoring. *Psychological Review*, *101*, 166–171.
Klauer, K. C., Musch, J., & Naumer, B. (2000). On belief bias in syllogistic reasoning. *Psychological Review*, *107*, 852–884.
Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747–1749.
Kokinov, B., & Petrov, A. (2001). Integrating memory and reasoning in analogy-making: The AMBR model. In D. Gentner, K. Holyoak, and B. Kokinov (Eds.), *The analogical mind: Perspectives from cognitive science* (pp. 59–124). Cambridge, MA: MIT Press.
Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.
Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual stimulus recognition. *Journal of Experimental Psychology: General*, *132*, 351–378.
Liew, J., Hayes, B. K., & Grisham, J. (2012). *Distortions of reasoning in obsessive-compulsive disorder*. Paper presented at 39th Annual Conference of the Australasian Experimental Psychology Society, Sydney.
Lamberts, K. (2002). Feature sampling in categorization and recognition of objects. *Quarterly Journal of Experimental Psychology: Section A*, *55*, 141–154.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–257.
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Macmillan, N. A., & Rotello, C. M. (2006). Deciding about decision models of remember and know judgments: A reply to Murdock. *Psychological Review*, *113*, 657–665.
McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508–520.
Markovits, H., & Nantel, G. (1989). The belief-bias effect in the production and evaluation of logical conclusions. *Memory & Cognition*, *17*, 11–17.
Medin, D. L., Coley, J. D., Storms, G., & Hayes, B. K. (2003). A relevance theory of induction. *Psychonomic Bulletin & Review*, *10*, 517–532.
Medin, D. L., Goldstone, R. L., & Markman, A. B. (1995). Comparison and choice: relations between similarity processing and decision processing. *Psychonomic Bulletin and Review*, *2*, 1–19.
Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 882–894.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
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.

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 Sciences,* USA. *109*, 333–338.
Nosofsky, R. M., & Zaki, S. R. (1998). Dissociations between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255.
Oaksford, M., & Chater, N. (2002). Common sense reasoning, logic and human rationality. In R. Elio (Ed.), *Common sense, reasoning and rationality* (pp. 174–214). Oxford: Oxford University Press.
Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford: Oxford University Press.
Oberauer, K., Süss, H., Wilhelm, O., & Wittmann, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641–652.
Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.
Pazzaglia, A., Dube, C., & Rotello, C. M. (2012). Multinomial processing tree models for recognition memory: A critical review. *Unpublished manuscript*.
Petty, R. E., & Cacioppo, J. T. (1984). The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, *46*, 69–81.
Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. *Psychological Review*, *97*, 285–308.
Rehder, B., & Burnett, R. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264–314.
Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science*, *12*, 129–134.
Roberts, M. J., & Sykes, E. D. A. (2003). Belief bias and relational reasoning. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *56*, 131–154.
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
Ross, B. H. (1996). Category learning as problem-solving. In Medin, D. L. (Ed.), (1996). *The psychology of learning and motivation, Vol. 35* (pp. 165–192).
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.
Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory and Cognition*, *28*, 907–922.
Rotello, C. M., & Heit, E. (2009). Modeling the effects of argument length and validity on inductive and deductive reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1317–1330.
Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, *111*, 588–616.
Rotello, C. M., Masson, M. E. J., & Verde, M. F. (2008). Type I error rates and power analyses for single-point sensitivity measures. *Perception & Psychophysics*, *70*, 389–401.
Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*, 657–661.
Shafto, P., Coley, J. D., & Baldwin, D. (2007). Effects of time pressure on context-sensitive property induction. *Psychonomic Bulletin & Review*, *14*, 890–894.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General*, *121*, 278–304.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231–280.
Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General*, *133*, 166–188.
Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 984–1002.
Smith, S. W., & Vela, E. (2001). Environmental context-dependent memory: A review and a meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, *117*, 34–50.
Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven: Yale.
Swets, J. A. (1986a). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*, 100–117.
Swets, J. A. (1986b). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, *99*, 181–198.
Townsend, C., & Heit, E. (2011). Judgments of learning and improvement. *Memory & Cognition*, *39*, 204–216.
Van Orden, G. C., Jansen op de Haar, M. A., & Bosman, A. M. T. (1997). Complex dynamic systems also predict dissociations, but they do not reduce to autonomous components. *Cognitive Neuropsychology*, *14*, 131–165.
Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, *108*, 394–401.
Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*, 716–732.
Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 125–154.
Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*, 616–641.
Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *20*, 1341–1354.
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, *46*, 441–517.