Relations Between Inductive Reasoning and Deductive Reasoning

Evan Heit
University of California, Merced

Caren M. Rotello
University of Massachusetts Amherst

One of the most important open questions in reasoning research is how inductive reasoning and deductive reasoning are related. In an effort to address this question, we applied methods and concepts from memory research. We used 2 experiments to examine the effects of logical validity and premise–conclusion similarity on evaluation of arguments. Experiment 1 showed 2 dissociations: For a common set of arguments, deduction judgments were more affected by validity, and induction judgments were more affected by similarity. Moreover, Experiment 2 showed that fast deduction judgments were like induction judgments—in terms of being more influenced by similarity and less influenced by validity, compared with slow deduction judgments. These novel results pose challenges for a 1-process account of reasoning and are interpreted in terms of a 2-process account of reasoning, which was implemented as a multidimensional signal detection model and applied to receiver operating characteristic data.

Keywords: reasoning, similarity, mathematical modeling

An important open question in reasoning research concerns the relation between induction and deduction. Typically, individual studies of reasoning have focused on only one one task, rather than examining how the two are connected (Heit, 2007). To make progress on this issue, we have borrowed concepts and methods from memory research, which has faced similar questions about the number of memory processes and how to model them. For example, a lively debate exists about whether recognition judgments can be accounted for by a single familiarity process or whether two processes are needed: heuristic familiarity and a more accurate recollective process (e.g., Rotello, Macmillan, & Reeder, 2004; Wixted & Stretch, 2004). That issue is often examined in the remember–know paradigm (Tulving, 1985), in which subjects make a recognition judgment then state whether they just know they have seen the item or actually remember it. Under the two-process view, "know" judgments depend more on familiarity, whereas "remember" judgments depend more on recollection. Under the one-process view, "remember" judgments reflect a stricter response criterion than "know."

Here, we treat the relation between induction and deduction as a psychological question rather than as a question of how to demarcate inductive problems versus deductive problems (e.g., Skyrms, 2000). Our empirical strategy is to ask people to judge either inductive strength or deductive validity for a common set of

arguments (Rips, 2001). This technique can highlight similarities or differences between induction and deduction that are not confounded by the use of different materials (Heit, 2007). It also allows us to compare two major classes of theories of reasoning. In broad terms, one-process accounts suggest that people apply the same reasoning abilities to problems of induction and deduction rather than drawing on different mechanisms for the two tasks (Harman, 1999). If the same mechanisms apply to induction and deduction, one possible distinction between the two tasks is that deduction judgments are like "remember" responses in requiring a stricter criterion for a positive response, because greater certainty is necessary (Rips, 2001; Skyrms, 2000).

According to two-process accounts (Evans, 2008; Stanovich, 2009), both heuristic and analytic processes contribute to reasoning, with each process potentially assessing an argument as strong or weak. We propose that induction judgments would be particularly influenced by quick heuristic processes that tap into associative information about context and similarity that does not necessarily make an argument logically valid. Deduction judgments would be more heavily influenced by slower analytic processes that encompass more deliberative, and typically more accurate, reasoning. Although two-process accounts have provided an explanatory framework for many results (e.g., content effects, individual differences), they typically have not been implemented and fitted to data.

In an effort to directly contrast one- and two-process accounts of reasoning, Rips (2001) reported response reversals within pairs of arguments: One argument was more likely to be judged inductively strong, and the other was more likely to be judged deductively valid. He concluded that this result was evidence against a one-process account, which predicts the same order of arguments in both conditions. Heit and Rotello (2005) extended Rips's study to examine subjects' sensitivity to valid versus invalid arguments. If deduction and induction use the same information, differing only in terms of response criterion, then sensitivity (d'), reflecting the difference in responses to valid and invalid arguments, should be

the same for deduction and induction. Instead, we found a higher d' for deduction (1.69) than for induction (0.86).

The present study comprises novel tests of a recently proposed two-process model of reasoning (Rotello & Heit, 2009), assuming that induction and deduction judgments both tap into underlying heuristic and analytic processes but in different proportions. We derived this model from results showing that deduction judgments were more sensitive to validity, that induction judgments were more sensitive to length of argument, and that reducing fluency (using a hard-to-read font) increased the role of analytic processing in induction judgments, making them more sensitive to validity. Here, for the first time, we compare the effects of similarity on induction and deduction; we also compare speeded deduction judgments with unspeeded deduction judgments. Similarity is a central construct in some theories of inductive reasoning (Osherson, Smith, Wilkie, Lopez, & Shafir, 1990; Sloman, 1993) and is a key predictor of inductive strength judgments (Heit & Feeney, 2005), so it is valuable to compare the role of similarity in induction versus deduction. Furthermore, generally speaking, speeded reasoning tasks seem to have great potential to help study the cognitive processing underlying reasoning, particularly when multiple processes may be involved. Surprisingly, there have been relatively few such studies (e.g., Evans & Curtis-Holmes, 2005; Shafto, Coley, & Baldwin, 2007).

To anticipate, in Experiment 1, we focused on premise–conclusion similarity (invalid arguments seem stronger when there is greater similarity between categories in the premise and the conclusion; for a review, see Hayes, Heit, & Swendsen, in press). We found two dissociations: Similarity had more impact on induction, which depended more on heuristic processing, and logical validity had more impact on deduction, which depended more on analytic processing. In Experiment 2, we found that fast deduction judgments were like induction judgments: They showed a greater influence of similarity and lesser influence of validity, suggesting that analytic processing was attenuated in speeded deduction. These results were all accounted for by an implemented two-process model but could not be explained by a one-process model.

Experiment 1

Method

Subjects. Sixty-six students from the University of California, Merced, were paid to participate. They were randomly assigned to one of two conditions: induction ($n = 32$) or deduction ($n = 34$).

Stimuli. There were 142 questions, comprising arguments about the following kinds of mammals: bears, cats, cows, dogs, goats, horses, lions, mice, rabbits, and sheep. An example (invalid) argument is listed below.

Horses have Property X

———————————

Cows have Property X

Subjects were instructed to treat Property X as a novel biological property. Stimuli were created as follows. The 10 kinds of mammals, plus the mammal category itself, composed a set of 11 categories. These 11 categories were rotated through the premise and conclusion positions to yield 121 arguments. Of these, 21 were valid: 10 arguments were based on category inclusion with mammal as the premise category, such as the example listed below.

Mammals have Property X

———————————

Cows have Property X

Furthermore, 11 arguments had identical premise and conclusion categories, such as the example listed below.

Cows have Property X

———————————

Cows have Property X

To increase the proportion of valid arguments, we presented the valid items twice so that there were 42 valid arguments (29.6%) out of 142 questions.[1] Note that for invalid arguments, the similarity between the premise and conclusion categories varied widely.[2]

Procedure. Subjects were first given instructions on the definition of strong or valid arguments. Following Rips (2001), subjects in the induction condition were told that strong arguments were those for which "Assuming the information above the line is true, this makes the sentence below the line *plausible*." Likewise, the deduction instructions 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."

The 142 arguments were presented one at a time, on a computer, in a different random order for each subject. In the induction condition, subjects pressed one of two keys to indicate "strong" or "not strong." In the deduction condition, subjects indicated "valid" or "not valid." Each binary decision was followed with a 1–5 confidence rating; higher numbers indicated greater confidence.

Results and Discussion

Three subjects from the induction condition were excluded because they gave the same response for virtually every question, or they made more "strong" responses to invalid than to valid arguments.

We first assessed the proportion of positive ("strong" or "valid") responses to valid and invalid arguments (see Table 1). For the deduction condition, the average proportions were 0.94 and 0.04, respectively. For the induction condition, the average proportions were 0.95 and 0.12, respectively. Subjects were more likely to reject invalid arguments in the deduction condition than in the induction condition, Welch's unequal-variance $t'(32.1) = 2.04$, $p < .05$, suggesting greater influence of validity for deduction. As

---

[1] Strictly speaking, inclusion arguments are enthymemes, because they rely on a hidden premise, such as that all cows are mammals (Calvillo & Revlin, 2005). For simplicity, we refer to both the identity and inclusion arguments as valid, although we report separate analyses for these two kinds of items.

[2] Similarity was derived from a multidimensional scaling solution for mammal categories (Rips, 1975). Euclidean distance was physically measured from the published scaling solution and was transformed to similarity using a negative exponential function (cf. Shepard, 1987).

Table 1
Response Proportions From Experiments 1 and 2

Experiment	Condition	Positive response rate to valid arguments	Positive response rates to invalid arguments	
			Low similarity	High similarity
1	Induction	0.95	0.08	0.14
	Deduction	0.94	0.04	0.05
2	Fast	0.83	0.16	0.24
	Slow	0.96	0.09	0.11

in previous experiments (Heit & Rotello, 2005, 2008; Rotello & Heit, 2009), d' was greater for deduction (3.31) than for induction (2.82),[3] which also suggests that deduction judgments were more affected by validity. However, use of d' requires that the underlying distributions are equal-variance Gaussian (Macmillan & Creelman, 2005; Rotello, Masson, & Verde, 2008), which has not been observed in reasoning tasks (Heit & Rotello, 2005; Rotello & Heit, 2009). Apparently simpler measures, such as the difference between acceptance rates for valid and invalid problems, also entail assumptions that are not supported by reasoning data (Dube, Rotello, & Heit, 2009).

A better approach is to consider the area under the receiver operating characteristic (ROC) curves, which plot positive responses to valid items against positive responses to invalid items, as a function of confidence. The left-most point on the ROC reflects the highest confidence positive judgments; points to the right include responses of decreasing confidence. When correct responses are more frequent, the ROC curve falls closer to the upper left corner, and the area under the curve is greater, reflecting greater sensitivity to the validity of the argument. Equal-variance Gaussian evidence distributions lead to symmetrically curved ROCs, which would justify the use of d'; distributions that justify the use of a difference between correct and error response rates lead to linear ROCs.

Figure 1 shows the ROCs for both conditions, which are curved and asymmetric. Thus, neither d' nor the difference between correct responses and errors accurately summarizes performance. Therefore, we instead turned to the area under the ROC, which was greater overall for deduction than for induction. This difference reached the level of statistical significance for identity problems in deduction versus induction ($z = 2.84$, $p < .01$; see Metz, 1998) but not so for inclusion problems ($z = 0.71$, $p > .4$), perhaps because subjects did not consistently treat these as valid (see Footnote 3).

The data in Table 1 suggest greater sensitivity to similarity for induction than for deduction, as predicted by a two-process account. (We assigned arguments to the low-similarity set or to high-similarity set on the basis of a median split, using the measures described in Footnote 2.) We calculated, for each subject, the difference in the positive response rate to low- and to high-similarity invalid arguments. As expected, subjects' difference scores were larger in the induction condition than in the deduction condition, $t(61) = 2.53$, $p < .02$, indicating that responses to invalid arguments were more influenced by similarity in induction than in deduction.

To investigate the effects of similarity at a finer level, we performed a multiple regression analysis predicting the proportion

of positive responses for the 121 unique arguments. The data were pooled across the deduction and induction conditions, yielding 242 data points to be predicted. The two main predictors were validity (coded as 0 or 1) and the similarity of the animal categories in the premise and conclusion. We added the following predictors: Validity × Similarity, Condition, Condition × Validity, Condition × Similarity, and Condition × Validity × Similarity. The overall R^2 was .979, $F(7, 234) = 763.5$, $p < .001$. Only three predictors were significantly different from zero. There were positive main effects of validity ($\beta = .93$, $p < .001$) and similarity ($\beta = .06$, $p < .01$). Crucially, there was a Condition × Similarity interaction ($\beta = .21$, $p < .001$), indicating more impact of similarity in the induction condition.[4]

Because we focused on similarity effects by systematically varying premise–conclusion similarity in this experiment, it is perhaps unsurprising that the most robust results were differences in the role of similarity for induction versus deduction. We also observed differences in sensitivity to validity, for example, subjects were significantly more likely to reject invalid arguments in the deduction condition than in the induction condition. These results add to previous findings of greater sensitivity to validity for deduction (Heit & Rotello, 2005, 2008; Rotello & Heit, 2009). We acknowledge that that limited range of the arguments in the present experiment, for example, each had just one premise, may have reduced the opportunity to find differences in sensitivity to validity (for related comments on using a wider range of arguments, see Oaksford & Hahn, 2007).

Modeling. Rotello and Heit (2009) found that subjects given induction instructions were heavily influenced by argument length (they were more likely to judge invalid arguments as strong when they included more premises); subjects given deduction instructions on the same arguments did not show that trend. We successfully fitted data from those three experiments using a two-dimensional signal detection model (Macmillan & Creelman, 2005) in which valid and invalid arguments differ along dimensions of consistency with prior knowledge (*heuristic evidence*) and apparent logical correctness (*analytic evidence*). Differences between the induction and deduction responses were explained by differential weighting of the two dimensions, which was reflected in the slope of the decision bound that divides "valid" from "invalid" or "strong" from "not strong" arguments. We were unable to fit the ROCs 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 argument length on both induction and deduction, and because it incorrectly predicted that accuracy would not differ between tasks.

[3] As in previous studies (e.g., Sloman, 1998), there was a higher proportion of positive responses to identity arguments (0.97 for deduction and 0.98 for induction) than to inclusion arguments (0.90 for deduction and 0.91 for induction). Identity arguments had a greater d' for deduction (3.64) than for induction (3.33). Inclusion arguments also had a greater d' for deduction (3.04) than for induction (2.53).

[4] Because more extreme probabilities have lower variance than probabilities near the middle of the range, we have also performed these regressions using the arcsine transformation recommended by Cohen, Cohen, Aiken, and West (2002). For both experiments, these regressions lead to the same conclusions.
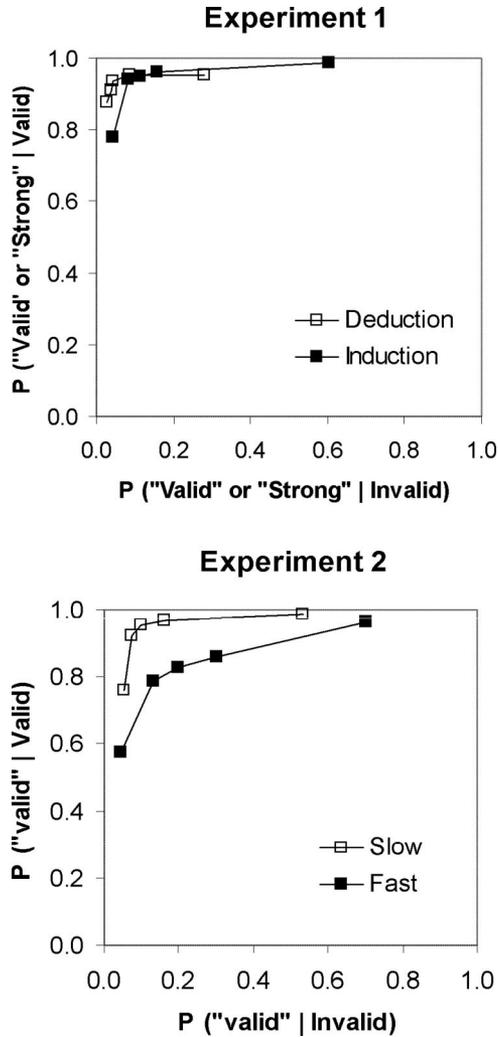
Experiment 1

Experiment 2

Figure 1. Receiver operating characteristic curves for Experiments 1 and 2.

A one-dimensional model cannot account for the present data, for analogous reasons: Induction and deduction resulted in different accuracy levels, and the effect of similarity was greater for induction than for deduction. Therefore, we focused on the two-dimensional model that Rotello and Heit (2009) found to be successful, assuming that consistency with prior knowledge could reflect the similarity of the premise category to the conclusion category (see Figure 2). Using Monte Carlo simulations, we sampled 1,000 trials each from distributions of high- and low-similarity invalid arguments (these were allowed to differ only in mean similarity) and from distributions of valid identity and inclusion arguments (these were allowed to differ in their mean values on both dimensions). Predicted hit and false-alarm rates were calculated from the proportion of sampled strengths that fell above the induction or deduction decision bound; simulated confidence level was manipulated by varying the y-intercepts of the decision bounds, generating ROCs. The parameters of the model (e.g., means, variances, and covariances of the distributions on each axis, decision bound slopes) were varied over a wide range, and a good fit was identified by considering both the mean squared

error (difference between observed and predicted response rates) and the area under the ROC, although the results should be considered illustrative rather than detailed quantitative fits. The resulting parameters are shown in Table 2.

Replicating Rotello and Heit's (2009) conclusions, the sole difference between induction and deduction in this model is the slope of the decision bound. Put differently, the modeling indicates that the only difference between induction and deduction is in terms of the relative weight that 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 3; Figure 4 shows the ROCs for the identity and inclusion arguments. Both figures show that the model fits the data well: The simulated ROCs fall within the 95% confidence intervals for the observed data. 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 (see Figure 3). Also, the predicted identity ROC falls higher in the space in the deduction condition than the induction condition (see Figure 4), because inclusion problems have a lower mean value on the validity dimension and that dimension is weighted more heavily in deduction.

To summarize, Experiment 1 shows that similarity has a greater effect on induction, and validity has a greater effect on deduction. These novel results are naturally accommodated by a two-process account of reasoning, but it is unclear how a one-process account would explain the results.

Experiment 2

Experiment 1 suggests that deduction judgments reflect a relatively greater influence of analytic processes than do induction
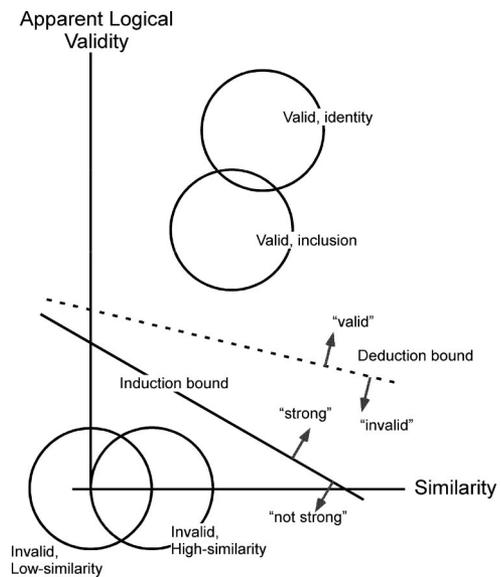

Figure 2. Schematic two-dimensional model for both induction and deduction. The circles show distributions of arguments in argument space, reflecting two underlying dimensions of argument strength. The dashed and solid lines show decision boundaries for deduction and induction, respectively.

Table 2
Parameter Values for the Two-Dimensional Model as Applied to Each Experiment

Parameter	Experiment 1	Experiment 2
d_x = mean of valid high-similarity arguments on x-axis	1.0	1.0
Variance of d_x	0.9	1.2
d_y = mean of valid high-similarity arguments on y-axis	3.7	3.5
Variance of d_y	1.5	1.6
Location of invalid high-similarity arguments on x-axis	1.0	0.2
Induction (or fast condition) slope	-0.5	-2.0
Deduction (or slow condition) slope	-0.1	-0.2
Change in d_x for inclusion arguments	0	-0.5
Change in d_y for inclusion arguments	-1.0	-0.8
Covariance of x and y for valid arguments	0	0
Covariance of x and y for invalid arguments	0	0

Note. The distribution of invalid low-similarity arguments was located at (0, 0) with a variance of 1 on each dimension and with no covariance.

judgments. In Experiment 2, we targeted the implications of a key assumption of two-process models, namely that analytic processes are relatively slow compared with heuristic processes. If deduction judgments are speeded up, they should tend to resemble induction judgments, because the contribution of analytic processes will be reduced. As far as we are aware, this prediction has never been previously tested. Subjects in a speeded condition were asked to make deduction judgments prior to a response deadline; another group was required to respond after a fixed delay. We expected to see a greater influence of similarity on the speeded responses (like the induction condition of Experiment 1), even though both groups were given deduction instructions. Our model-based prediction was that the slope of the decision bound would be steeper in the speeded condition, reflecting greater weighting of the similarity information and reduced weighting of the analytic information.

Method

The method was the same as Experiment 1, except for the following: 62 individuals participated (slow, $n = 34$; fast, $n = 28$).

All subjects received deduction instructions. Subjects were provided with five practice problems on unrelated materials to learn the timing of the task. In the slow condition, subjects were required to wait 8 s before making a response. The computer did not allow any input until 8 s had elapsed. In the fast condition, subjects were instructed to respond as quickly as possible and within 3 s. After 3 s, the display was cleared, and the subject was not allowed to respond to that argument.

Results and Discussion

Two subjects were excluded from the data analyses (one in each condition), according to the same criteria as Experiment 1. For the
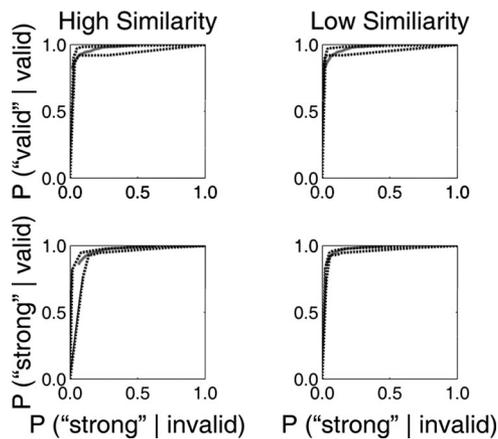
Figure 3. Simulated receiver operating characteristics (solid functions) generated with the two-dimensional model sketched in Figure 2 and the parameter values shown in Table 1, as well as 95% confidence intervals for the observed receiver operating characteristics (dashed functions) from Experiment 1. Top row: deduction condition; bottom row: induction condition; left column: high-similarity invalid arguments; right column: low-similarity invalid arguments.
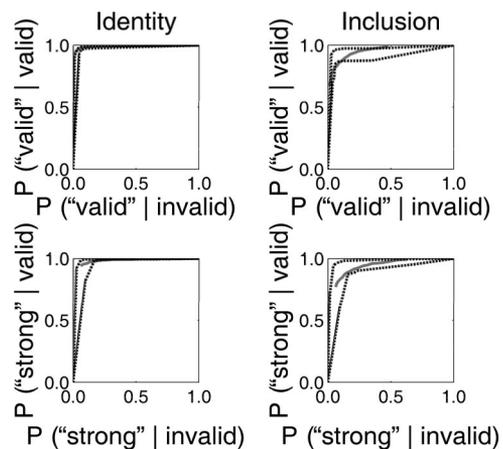
Figure 4. Simulated receiver operating characteristics (solid functions) generated with the two-dimensional model sketched in Figure 2 and the parameter values shown in Table 1, as well as 95% confidence intervals for the observed receiver operating characteristics (dashed functions) from Experiment 1. Top row: deduction condition; bottom row: induction condition; left column: identity arguments; right column: inclusion arguments.

remaining subjects in the fast condition, an average of 2.2% of responses was missing because of slow responding.

We first assessed the proportion of positive responses to valid and invalid arguments (see Table 1). For the slow condition, the average proportions were 0.96 and 0.10, respectively. For the fast condition, the average proportions were 0.83 and 0.20, respectively. Analogously to the deduction and induction conditions in Experiment 1, subjects were more likely to reject invalid arguments in the slow condition than in the fast condition, $t(58) = 2.15$, $p < .05$, suggesting greater sensitivity to validity for the slow condition. Moreover, d' was greater for the slow condition (3.03) than for the fast condition (1.80).[5] The ROCs, in the lower panel of Figure 1, are once again asymmetric and curved; they clearly show that faster responses led to lower sensitivity to the distinction between valid and invalid arguments ($z = 11.70$, $p < .001$).

We predicted that similarity would have a greater influence in the fast condition than in the slow condition. As in Experiment 1, we tested this hypothesis by calculating the difference between the positive response rates to low- and to high-similarity invalid arguments. Subjects' responses to invalid arguments were more influenced by similarity in the fast condition than in the slow condition—that is, difference scores were larger, $t(58) = 2.74$, $p < .01$.

Another way of looking at the effects of similarity is through multiple regression. Using the same predictors as in Experiment 1, we calculated that the overall R^2 was .927, $F(7, 234) = 422.1$, $p < .001$. Only four predictors were significantly different from zero. There were positive main effects of validity ($\beta = .57$, $p < .001$) and similarity ($\beta = .20$, $p < .001$). Crucially, there was a Condition × Similarity interaction ($\beta = .28$, $p < .001$), indicating more impact of similarity in the fast condition. There was also a Condition × Validity interaction ($\beta = .33$, $p < .01$), indicating more impact of validity in the slow condition.

We used Monte Carlo simulations to evaluate a two-dimensional model of the differences between the fast and slow conditions, on the assumption that the fast deduction condition would appear more like an induction task (having a steeper decision bound and greater weighting of similarity information). The resulting ROCs for high- and low-similarity arguments are shown in Figure 5; Figure 6 presents the ROCs for the inclusion and identity arguments. The model captures the data well by varying only the slope of the decision bound between fast and slow judgments (see Table 2). In Figure 5, the ROC shifts to the right for high similarity compared with low similarity, and for fast judgments but not for slow judgments, reflecting a greater similarity effect. Likewise, Figure 6 shows that the ROCs for identity problems fall higher in the space than those for inclusion problems and that the ROCs fall higher in space for slow judgments than for fast judgments.

To summarize, Experiment 2 shows two dissociations between fast and slow deduction judgments: Fast judgments were influenced more by similarity and slow judgments more by validity. These results parallel the findings of Evans and Curtis-Holmes (2005), who observed that fast judgments on categorical syllogisms were more influenced by prior beliefs, and slow judgments were more influenced by validity. Our speeded deduction results join Shafto et al.'s (2007) data on speeded induction judgments in showing a greater impact of taxonomic similarity on speeded responses compared with slower responses. Such results are natural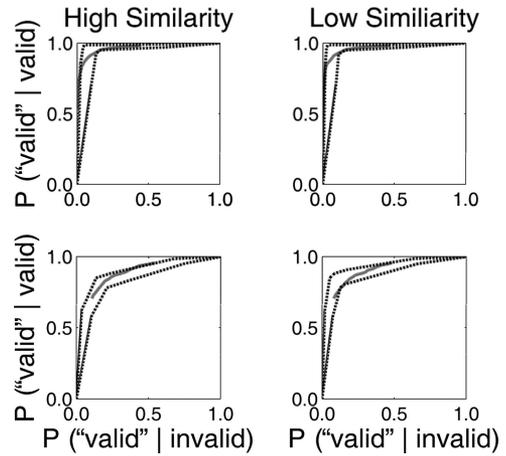ly accounted for by two-process accounts of reasoning, assuming that fast judgments are dominated by heuristic processes and that analytic processes contribute more heavily to slow judgments.

Two-process accounts themselves have different varieties (Evans, 2008). For example, analytic processes might follow heuristic processes sequentially or these processes might run in parallel. Although the finding of reduced analytic processing under speeded conditions is compatible with the sequential view, it can also be explained by the parallel view with the assumption that analytic processing runs more slowly. The present analyses do not distinguish between these particular alternatives, but more generally, we expect that multidimensional signal detection models will be helpful in constraining future process models of reasoning.

General Discussion

There is much to gain by building bridges between the psychology of memory and the psychology of reasoning; mathematical modeling provides one such overarching structure. We have made progress on the important question of how inductive reasoning and deductive reasoning are related. A growing body of evidence now shows that asking people to make either deduction judgments or induction judgments draws on somewhat different cognitive resources, even when people are judging exactly the same arguments (for a review of related evidence from brain imaging research, see Hayes et al., in press). Our Experiment 1 highlights differences between induction and deduction in terms of the influence of similarity. Experiment 2 demonstrates that fast deduction judgments are like induction judgments in terms of being affected more

Figure 5. Simulated receiver operating characteristics (solid functions) generated with the two-dimensional model sketched in Figure 2 and the parameter values shown in Table 1, as well as 95% confidence intervals for the observed receiver operating characteristics (dashed functions) from Experiment 2. Top row: slow condition; bottom row: fast condition; left column: high-similarity invalid arguments; right column: low-similarity invalid arguments.

---

[5] As in Experiment 1, there was a higher proportion of positive responses to identity arguments (0.98 for slow and 0.95 for fast) than to inclusion arguments (0.93 for slow and 0.69 for fast). Here, identity arguments had a d' of 3.25 for slow and 2.50 for fast. Inclusion arguments had a d' of 2.79 for slow and 1.34 for fast.
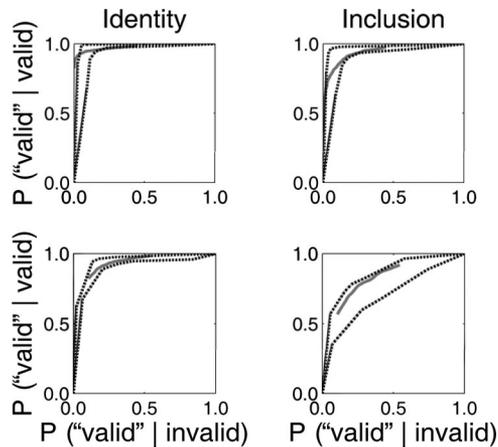
Figure 6. Simulated receiver operating characteristics (solid functions) generated with the two-dimensional model sketched in Figure 2 and the parameter values shown in Table 1, as well as 95% confidence intervals for the observed receiver operating characteristics (dashed functions) from Experiment 1. Top row: slow condition; bottom row: fast condition; left column: identity arguments; right column: inclusion arguments.

by similarity and less by logical validity. Simulations show that the results of both experiments are consistent with a two-process model of reasoning in which deduction and induction judgments result from different weighting of the information from two underlying processes. That is, we were able to model the difference between deduction and induction, and between unspeeded and speeded deduction, the same way.

What are the implications of these results for other models of reasoning in the literature? Some very successful accounts (Johnson-Laird, 1994; Oaksford & Chater, 2002; Osherson et al., 1990; Sloman, 1993) have applied a common modeling framework to both inductive and deductive arguments, with a single scale of evidence for argument strength.[6] Hence, these accounts bear some similarity to the idealized one-process account investigated by Rips (2001), but of course each model has additional mechanisms. Although these accounts can distinguish between different kinds of arguments (e.g., inductively strong vs. deductively valid), these accounts have not made explicit predictions for different kinds of judgments (induction vs. deduction). Further assumptions would be needed to extend these models to make two different kinds of judgments (and speeded vs. unspeeded judgments). Also, these models do not maintain separate sources of information about validity and similarity. In contrast, in memory research, it has been suggested that a one-dimensional signal detection model could successfully account for results pointing to two processes in recognition memory by combining information about recollection and familiarity into a single scale (Wixted & Stretch, 2004). Perhaps other models of reasoning could be developed along these lines, combining validity and similarity information into a single scale. However, other current models of reasoning do not seem to work that way now. For example, Bayesian models (Oaksford & Chater, 2002; see also Heit, 1998; Tenenbaum & Griffiths, 2001) measure the strength of an argument in terms of the probability of its conclusion, and they do not have component probabilities for validity and similarity. In conclusion, we would not rule out future

versions of other models of reasoning, on the basis of the present results as well as those in Rotello and Heit (2009). Indeed, we hope that other models will be developed further to address these challenging results.

Our own efforts have targeted the detailed predictions of two-process models of reasoning, including implementing these models. Here, and in Rotello and Heit (2009), we have demonstrated that induction and deduction can be described as relying on different weighted functions of the same information (i.e., different decision bounds in a common representation). Thus, our modeling supports a two-process model of reasoning; our data would be difficult for a one-process model to accommodate. Although our models represent explicit hypotheses about how two sources of evidence are brought to bear on making judgments, we do not claim to have developed a process model of reasoning, although the multidimensional models we have developed could constrain process models. Two-process theories of reasoning themselves can make a variety of mechanistic assumptions (Evans, 2008; Stanovich, 2009). For example, our results are compatible with a more automatic, similarity-based mechanism as well as what Evans (2008) calls an *intervention* and Stanovich (2009) calls an *override*—given enough time, people have the potential to substitute optional analytic processing that is sensitive to logical validity for their automatic, similarity-based processing. More generally, we suggest that by varying instructions and time to respond, there will be a rich set of results in reasoning research that will be an important test bed for developing and testing models of reasoning.

---

[6] Johnson-Laird (1994) explained how mental model theory, typically applied to problems of deduction, can also be applied to problems of induction. Although allowing that people might explicitly perform deductive tasks under limited circumstances, Oaksford and Chater (2002) showed that a probabilistic reasoning model, extending logic to a probabilistic scale of validity, can be applied to problems of deduction. Osherson et al. (1990) and Sloman (1993) presented models of induction that account for some deductive reasoning phenomena (e.g., that arguments based on identity matches between a premise and a conclusion are perfectly strong).

References

Calvillo, D. P., & Revlin, R. (2005). The role of similarity in deductive categorical inference. *Psychonomic Bulletin & Review, 12,* 938–944.

Cohen, J., Cohen, P., Aiken, L. S., & West, S. G. (2002). *Applied multiple regression—Correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Dube, C., Rotello, C. M., & Heit, E. (2009). *Assessing the belief bias effect with ROCs: It's a response bias effect.* Manuscript submitted for publication.

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

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

Harman, G. (1999). *Reasoning, meaning, and mind.* Oxford, England: Oxford University Press.

Hayes, B. K., Heit, E., & Swendsen, H. (in press). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science.*

Heit, E. (1998). A Bayesian analysis of some forms of inductive reasoning. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 248–274). Oxford, England: Oxford University Press.

Heit, E. (2007). What is induction and why study it? In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 1–24). Cambridge, England: Cambridge University Press.

Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review, 12,* 340–344.

Heit, E., & Rotello, C. M. (2005). Are there two kinds of reasoning? In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 923–928). Mahwah, NJ: Erlbaum.

Heit, E., & Rotello, C. M. (2008). Modeling two kinds of reasoning. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 1831–1836). Austin, TX: Cognitive Science Society.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50,* 189–209.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Metz, C. E. (1998). ROCKIT [Computer software]. Chicago, IL: Department of Radiology, University of Chicago. Retrieved from http://xray.bsd.uchicago.edu/krl/roc_soft6.htm

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, England: Oxford University Press.

Oaksford, M., & Hahn, U. (2007). Induction, deduction, and argument strength in human reasoning and argumentation. In A. Feeney & E. Heit (Eds.), *Inductive reasoning* (pp. 269–301). Cambridge, England: Cambridge University Press.

Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 97,* 185–200.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

Rips, L. J. (2001). Two kinds of reasoning. *Psychological Science, 12,* 129–134.

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.

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, September 11). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.

Skyrms, B. (2000). *Choice and chance: An introduction to inductive logic* (4th ed.). Belmont, CA: Wadsworth.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology, 25,* 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology, 35,* 1–33.

Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought.* New Haven, CT: Yale.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24,* 629–641.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11,* 616–641.

Correction to Son (2010)

In the article "Metacognitive Control and the Spacing Effect," by Lisa K. Son (*Journal of Experimental Psychology: Learning, Memory, and Cognition,* 2010, Vol. 36, No. 1, pp. 255–262), lenient scores were reported instead of strict scores in two Performance Data sections of the text. The strict scores were correctly used in the analyses and figures.

On page 259, the data corrections are as follows: The mean level of performance for items that were massed was 17.3 rather than 27.48, whereas that of spaced items was 30.6 rather than 34.02. The mean performance for those items in which a spacing schedule was imposed was 22.6 rather than 28.90, and the mean for the massed items was 21.9 rather than 27.48.

On page 260, the data corrections are as follows: The mean for the massed items was 5.0 rather than 10.3; for spaced items, the mean was 29.3 rather than 36.2. Children using the forced spacing strategy had a mean performance of 11.7 rather than 20.7. This mean score was still almost double that of the forcibly massed items, $M = 5.2$ rather than 11.1.