Title
Reasoning with differing tasks and response formats

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Cruz de Echeverria Loebell, Nicole
Knauff, Markus

Publication Date
2013

Peer reviewed

Reasoning with differing tasks and response formats

Nicole Cruz de Echeverría Loebell (Nicole.Cruz@psychol.uni-giessen.de)[1]
Markus Knauff (Markus.Knauff@psychol.uni-giessen.de)[1]

[1]Justus Liebig University Gießen, Experimental Psychology and Cognitive Science,
Otto-Behaghel-Str. 10F, Giessen, 35394 Germany

Abstract

This study investigated the role probabilistic and deductive relations play in the reasoning process. It was predicted that when taking an analytic stance to a problem, it would take longer to evaluate inferences when asked how probable it is that the conclusion is true, than when asked whether the conclusion follows or not from the premises. Contrary to this prediction, people responded faster when the response format was continuous. However, there was no effect of argument type with continuous response format, suggesting people did not assess entailment relations in this condition. Options to address the issue further are discussed.

Keywords: deductive/inductive reasoning; dual process theories; task effects; response times.

Introduction

"If the animal is a whale, then it must be a mammal"; "If I stay for five more minutes, I shall still catch the train"; "If you exchange these two cables, the telephone will work again". We go about the world constantly making judgments about what might be the case and what consequences we may expect from different situations and actions. Sometimes the reasoning involved occurs rather automatically, at other times it is effortful and time consuming. A lot of it involves conditionals, i. e. statements of the form "if p then q", with "p" and "q" standing for individual propositions such as "you stay 5 more minutes" and "you catch the train".

There is a debate in reasoning research regarding what criteria, or norms for when an inference is correct, people employ when drawing inferences – and, if they employ different criteria, then under what circumstances they reason according to what criterion and in what way the criteria may interact. The two main norms under discussion are a deductive, deterministic one (the conclusion is correct if it follows necessarily from the premises) and a probabilistic one (e. g. the conclusion is correct if its uncertainty is not greater than the sum of the uncertainties of the premises, Adams, 1975). People's answers to reasoning problems are generally sensitive both to the structure of deductive entailment relations involved and to the subjective probability or plausibility of the contents appearing in the relations (i. e. Thompson, 1994; Singmann & Klauer, 2011).

In some approaches it is argued that people reason using a single norm for argument validity across situations. A major proponent of this position, the theory of mental models (Johnson-Laird & Byrne, 2002), postulates this to be the deductive norm and accounts for the effect of contextual and probabilistic information on people's inferences by proposing that people integrate such information in their models of the situation, either by adding or subtracting possibilities considered, or by tagging the models with probabilities (Girotto & Johnson-Laird, 2004; Johnson-Laird, Legrenzi, Legrenzi, Girotto, & Caverni, 1999). A further major proponent of the single-criterion position is the probabilistic theory of Oaksford and Chater (Oaksford, Chater, & Larkin, 2000; Oaksford & Chater, 2007), which postulates that the effect of contextual and probabilistic information is a consequence of that people generally reason not deductively but probabilistically, in a way that is ecologically rational and that can be modeled using Bayesian theory together with a few further assumptions.

The idea that people use a single norm for argument validity across situations is put into question by a number of findings. Rips (2001) found that when given the same list of arguments which were valid/invalid as well as plausible/implausible, a group of people given deductive instructions endorsed the valid but implausible arguments more often than the invalid but plausible ones. The opposite was the case for a group of people given inductive instructions. Vadeboncoeur and Markovits (1999) found that emphasizing the deductive nature of a task in the instructions led to answers in stronger accordance with such instructions, but that also then the availability of counterexamples to the arguments (making them less plausible even though they were valid) still had an effect. Also the availability of probabilistic information was found to have an effect on people's approach to reasoning problems. For instance, Wolf and Knauff (2008) found that people's strategy of belief revision with conditional inferences was a function of the probability of the conditional when this probability was high or low, but was better explained by the theory of mental models when the probability of the conditional was close to .5 and thus perhaps less informative. A further factor found to influence people's reasoning is the task employed. For example, across several studies the theory of mental models offered a better explanation of reasoning in the conditional inference task, while the probabilistic approach could explain better findings in the truth table task, which is related more directly to the interpretation of conditionals (Geiger & Oberauer, 2010). Finally, also the response format for otherwise identical tasks, especially whether this is dichotomous or not, has been found to play a role. Oberauer, Geiger, Fischer, and Weidenfeld (2007) found that in the truth table task, the same participants who answered in accordance with a probabilistic interpretation of the conditional having the three response options "true", "false", and "irrelevant", answered in accordance with a mental model interpretation when the response option "irrelevant" was not available. Further, Markovits and Handley (2005) found that while probability ratings of the arguments of the conditional inference task where uniformly high, proportion of endorsement of the same inferences having binary response format was significantly lower, especially when the inferences where deductively invalid.

Findings like the ones described have led to increasing attempts to find integrative approaches, often in the form of dual-process theories, which assume that people may employ different criteria and ways of thinking under different circumstances. Hereby one process is often described as analytic, under more conscious control, more dependent on working memory resources and more context independent, and the other as heuristic, fast, automatic, context dependent and not much affected by working memory constraints. For instance, Klauer, Beller, & Hütter (2010) distinguish between a process based on the "logical form" or entailment relations in an argument and one based on content and context information. Sloman (1996, 2002) distinguishes between an associative and a rule based process. The two processes can be related in different ways. For instance, Evans and Over (Evans, 2006; Evans, Handley, Neilens, & Over, 2010) advocate a default-interventionist relation, in which the heuristic process is used as the default, and the analytic process may intervene if there is enough time and the heuristic answer seems insufficient to solve the task. Verschueren, Schaeken, and d'Ydewalle (2005) propose that both processes operate in parallel on a given task, and if the analytic process has enough time and leads to a different result than the heuristic process, it will override the answer arrived at by the heuristic process.

One difficulty with dual-process theories is that they often only explain the effect of deductive validity through the analytic system, while the construction of a representation of the problem to be evaluated can be better attributed to the heuristic system. This puts into question their role as independent forms of solving the same reasoning problem. Also, findings from de Neys (e. g. 2012) suggesting people have not only intuitive heuristics but also logical intuitions, question the idea of an association between the heuristic and the probabilistic on the one hand, and the analytic and the deductive on the other.

The present study aims at investigating further the role of deductive and probabilistic aspects of the reasoning process. Although in general it is plausible that people may approach a task in different ways depending on their goals and constraints of the situation, it is hypothesized that at least some of the findings proposed as evidence for two systems of reasoning may also be explained by making a less strong assumption: through the idea that the reasoning process is a composite one, in which different processes take over different components of the reasoning task, instead of reflecting different approaches to the same task. The two components considered here are assessment of the probability that a statement is the case (related to the interpretation of the statement) and assessment of what follows from the assumption that a statement is the case. The task of assessing whether something is the case is considered probabilistic: in the context of a conversation, it would be a matter of debate and subject to varying degrees of confidence. In contrast, the task of assessing what follows from the assumption that something is the case is considered (given a deductive task) as deductive and thus in a way deterministic, not probabilistic (something follows or it does not follow from given assumptions). In daily life we are often interested not just in what follows from assuming a certain piece of information, but also in how probable the conclusion itself is: we want to take into account also the uncertainty in the premises and transfer it to the conclusion. However, this is proposed to be a separate task within the reasoning process.

Thus, we hypothesized that, provided people approach a task analytically, it should take longer to answer to the question: "how probable is it that the conclusion from the premises is true?" than to the question: "does the conclusion follow from the premises?" Conversely, if people are given not inferences but only statements to evaluate, it should be faster to answer to the question: "how probable is it that this statement is true?" than to the question: "is this statement true or false?" since the latter case would involve the additional task of setting a threshold - above which one says "yes, it is true" and below which one says it is false - and of comparing the probability of the statement with this threshold. In order to raise the probability that people approach the task analytically, people are often given no time pressure as well as deductive instructions emphasizing the importance of assuming the truth of the premises for the sake of argument. We gave participants no time pressure, but could not emphasize deductive instructions since we wanted to assess the effect of taking into account premise probabilities in addition to entailment relations. We hoped that enough participants would nonetheless take an analytic stance given that in dual-process theories the weight obtained for the parameter representing an analytic approach to the task was often above 50% for both binary (Oberauer, 2006) and continuous (Klauer, Beller, & Hütter, 2010) response formats.

Method

Participants

Thirty-two students from the University of Giessen took part in the experiment in exchange for payment or course credit. Their mean age was 23.6 years (range: 19-31). They came from different majors, with the exclusion of mathematics, informatics, physics and philosophy. One participant[1] had taken a course in logic; sixteen had taken at least one course in statistics.

[1] This participant did not show a deterministic response pattern, and her exclusion did not change the pattern of results.

Design

The above hypotheses were assessed through a within subject design involving the two main variables *task* (evaluation of statements or of inferences) and *response format* (continuous, dichotomous). For statements, a further distinction was made between conditional statements ("if p then q") and the two statements the conditional is composed of ("p" and "q"). For inferences, one could further distinguish inference form. There were four inference forms: "Modus Ponens" (MP: "if p then q", "p", therefore "q"), "Modus Tollens" (MT: "if p then q", "not-q", therefore "not-p"), "Affirmation of the consequent" (AC: "if p then q", "q", therefore "p"), and "Denial of the antecedent" (DA: "if p then q", "not-p", therefore "not q"). Only the first two are deductively valid, because in the other two cases also the negation of the conclusion is compatible with the premises (However, if the conditional is interpreted as a biconditional: "p if and only if q" then all four inferences are deductively valid). The main dependent variable was response latency, but degree of resp. frequency of endorsement was also examined.

Material and procedure

Participants viewed either statements or inferences on the computer screen, and were asked to evaluate them on a continuous or dichotomous scale. Statements and inferences were embedded in one of four contexts involving concrete materials but describing arbitrary relations. For example, one such context was the following:

> In a workshop in Soko there is a cupboard with blue and yellow drawers for storing the nails and screws. One drawer of the cupboard is opened...

On the next screen appeared the statement or inference to be evaluated, e. g. "If the drawer is blue, then there are nails in it". There were three types of statements: conditionals like the one above (p -> q), and two statements corresponding to the antecedent (p, e. g. "the drawer is blue") and to the consequent (q, e. g. "the drawer has nails in it") of the conditional, respectively. There were four kinds of inferences, corresponding to MP, MT, AC and DA. For statements, participants were asked "How probable is it that this statement is true?" with continuous response format (cont), and "Is this statement true or false?" with dichotomous response format (dic). For inferences, the task was to "Consider the statements. How probable is it that the conclusion is true?" with continuous response format and "Assume the statements are true. Does the conclusion follow necessarily from them?" with dichotomous response format. Here we spoke of an evaluation of "the conclusion" and not of a specific statement per se, to make explicit that both response formats involve the evaluation of inferences and not just of statements grouped with other statements.

The continuous response scale was a horizontal line with the endpoints "0%" and "100%" and was divided into 101 points that could be clicked with the mouse. The dichotomous response scale consisted of two adjacent boxes, together as long as the horizontal line of the continuous response scale, below which stood the words "false" and "true" for statements, and "does not follow" and "follows" for inferences. To the right of each statement and each premise stood a small box filled up to a certain point, representing the probability of the statement (the fuller the box, the more probable the statement). There were four boxes representing the probabilities .2, .4, .6 and .8. The aim of these boxes was to provide premise probabilities in a non-numeric and yet relative standardized way.

Each of the four contexts was associated with the three statement types, yielding 12 statements for each response format. Further, each context was associated with the four inference types, leading to 16 inferences for each response format. For each participant, one of the four probabilities was randomly assigned to the conditional of one of the four contexts and held constant across the experiment, mimicking the reliability of conditional relations. For each context, the other three probabilities were distributed randomly without replacement across statements, such that e. g. for the context of the workshop, the second premise had a different probability for each of the four inferences. The order of occurrence of the statements and of the inferences was varied randomly for each participant.

Participants were tested individually in two sessions. One session involved evaluation of the 24 statements, the other evaluation of the 32 inferences. The order of sessions was counterbalanced across participants. Within each session, response format was blocked. Instructions at the beginning of each block included familiarization with the response scale and a sample trial. At the end of the second session, all participants worked through 20 trials in which the two response scales were presented alone on the screen (10 times each in random order) and they were to click with the mouse on them as quickly and as randomly as they could. This served to assess differences in response time to the two scales due to processes unrelated to the reasoning task (i. e. motor affordances). This difference was later subtracted from the answers to the reasoning task by centering the values of each participant in each response format around their mean for that response format when presented alone. The experiment was self-paced and lasted about 50 minutes.

Results and discussion

The data were analyzed separately for response times and for endorsement ratings as dependent variable. Prior to the analysis of response times, responses faster than 100 ms were eliminated, leading to exclusion of two data points. Elimination of response times outside the interval of the mean plus minus 3 SD for each variable led to no further data exclusions. Since response times have a lower threshold, they do not follow a normal distribution. To compensate for this, the inverse of response times: speed (1/RT), was taken for analysis. This normalizes somewhat the distribution and reduces the impact of outliers while preserving power and ease of interpretation (Whelan, 2008). Measures of speed were then multiplied by 1000 to avoid working with only very small values (Baayen & Milin, 2010). Prior to the analysis of endorsement ratings, it was necessary to represent the probability ratings obtained with continuous response format, and the endorsement frequencies obtained with dichotomous response format on the same scale. This was done by transforming mean frequency of the dichotomous items into a percentage value. For example, if a person answered three times yes (coded 1) and one time no (coded 0), the mean frequency of acceptance was (1 + 1 + 1 + 0)/4 = .75 = 75% (Markovits & Handley, 2005). It is thereby important to keep in mind that probability ratings and endorsement frequencies are different measures and may not be directly comparable. Results from such comparisons can be illustrative and useful, but should be interpreted with caution (Singmann & Klauer, 1010).

Separately for both response speed and endorsement ratings, three ANOVAS were conducted: a general ANOVA across tasks, assessing the effects of task (statements, inferences) and of response format (continuous, dichotomous); an ANOVA for statements assessing the effect of statement type (p -> q, p, q) and response format (cont, dic); and an ANOVA for inferences assessing the effect of inference type (MP, MT, AC, DA) and response format. The Greenhouse-Geisser correction of degrees of freedom for lack of sphericity was applied when appropriate. The results are depicted in Figure 1.

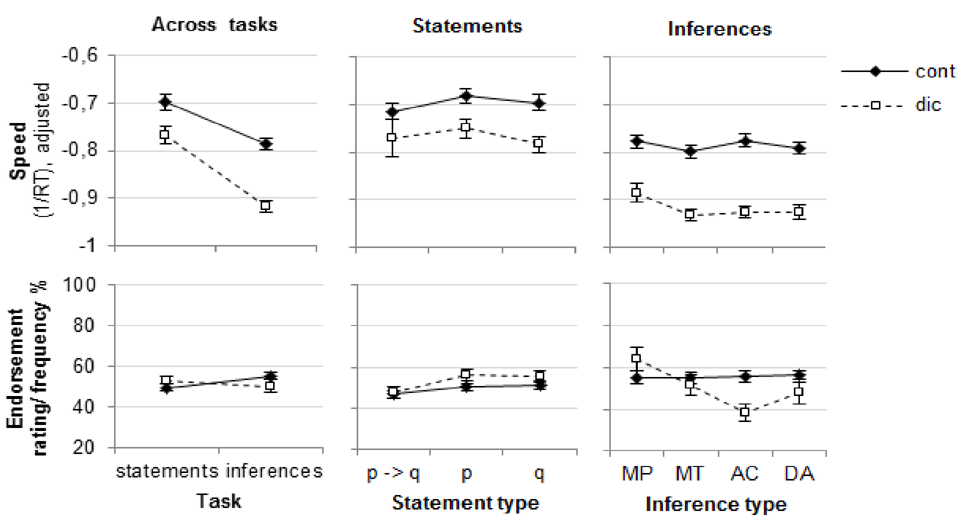

Figure 1. The upper panel shows mean speed (adjusted for RT differences between scales when presented alone) of responses for continuous (cont) and dichotomous (dic) response format, across tasks (left column), for statements (middle column) and for inferences (right column). The lower panel shows probability ratings (when response format = cont) resp. endorsement frequency (when response format = dic) for the same conditions. Error bars show within subject standard errors (Bakeman & McArthur, 1996).

For the sake of exposition clarity, only results considered relevant for the hypotheses will be reported in detail. The main hypothesis concerns the effect of task and of response format on response speed. Initially, this analysis was conducted using response speed not adjusted for differences in speed due to the scales alone. This analysis (not represented in Figure 1) yielded a main effect of task, $F(1, 31) = 134.72$, $p < .001$, partial $\eta^2 = .81$: answers to statements were faster than to inferences; a main effect of response format, $F(1, 31) = 14.96$, $p = .001$, partial $\eta^2 = .33$: answers were faster when the response format was dichotomous than when it was continuous; and an interaction between task and response format, $F(1, 31) = 6.58$, $p = .015$, partial $\eta^2 = .18$: the extent to which answers were faster when the response format was dichotomous was greater when evaluating statements than when evaluating inferences. This same ANOVA was then repeated correcting for differences in speed due to the scales alone, i. e. centering the values of each participant in each response format around the participant mean for that response format when presenting the scale alone. This analysis is shown in the upper left panel of Figure 1. It yielded a main effect of task, $F(1, 31) = 134.72$, $p < .001$, partial $\eta^2 = .81$: answers to statements were faster than to inferences; a main effect of response format, $F(1, 31) = 15.1$, $p = .001$, partial $\eta^2 = .33$: answers were faster when the response format was continuous; and an interaction between task and response format, $F(1, 31) = 6.58$, $p = .015$, partial $\eta^2 = .18$: the extent to which answers were faster when the response format was continuous was greater when evaluating inferences than when evaluating statements.

Thus, while in absolute terms it took longer to answer to the continuous than to the dichotomous scale, this relation was reversed when adjusting for differences in response times to each scale when presented alone, such that participants were faster when the response format was continuous. This is in accordance with our hypothesis for judgments about statements, but contrary to our hypothesis for judgments about inferences.

A possible explanation for why responses where faster with continuous response format both when evaluating statements and when evaluating inferences lies in the lower right panel of Figure 1, depicting endorsement ratings resp. endorsement frequency of the four inferences (MP, MT, AC, DA) as a function of response format. This analysis yielded no effect of response format, $F(1, 31) = 1.07$, $p = .31$, partial $\eta^2 = .03$; an effect of inference type, $F(3, 93) = 4.26$, $p = .007$, partial $\eta^2 = .12$; and an interaction between inference type and response format, $F(1, 31) = 5.77$, $p = .001$, partial $\eta^2 = .16$. The graphic shows the typically observed pattern of response to the four inferences for dichotomous response format (Bonferroni corrected *t*-tests only yielded a significant difference between MP and AC ratings, $t(31) = 4.3$, $p < .001$), whereas there was not a trace of an effect of inference type for continuous response format. Thus, people seem to have taken into account differences in the entailment relations making up the structure of the arguments only when the response format was dichotomous, but not when it was continuous. This finding renders it understandable that people were faster when the response format was continuous.

Finally, it is interesting to note that there was no effect of response format in all three analyses of endorsement ratings resp. endorsement frequency (lower three panels of Figure 1): In the ANOVA across tasks: $F < 1$; In the ANOVA for statements: $F(1, 31) = 3.44$, $p = .07$, partial $\eta^2 = .1$; and in the ANOVA for inferences: $F(1, 31) = 1.07$, $p = .31$, partial $\eta^2 = .03$.

The absence of an effect of response format in all three analyses speaks against the idea that people build a threshold close to certainty in the condition with binary response format, as had been suggested by Markovits and Handley (2005), who also compared answers in the conditional inference task with binary and continuous response format and found lower levels of inference endorsement when the response format was binary. It rather suggests people endorsed a probabilistic interpretation of the statements throughout: No effect of response format is expected when people judge a statement as true when they judge its probability to be above 50% and as false when they judge its probability to be below this value. This is a sensible strategy from a probabilistic perspective because then one's judgments will be right over 50% of the time on average. One explanation for the difference between our results and those of Markovits and Handley is that in our experiment one could explicitly see a representation of the statements' probabilities, and this may have made it more likely that they were taken into account as criteria for the judgments. One could assess the issue further using other means of providing probability information, such as through the introduction of a probability learning phase to simulate natural sampling, or through the use of familiar relations for which people can readily build probability estimates.

Although the absence of an effect of inference type for judgments with continuous response format provides a reason for why people's answers were generally faster when the response format was continuous, this absence of an effect is itself surprising and therefore worthy of further consideration. In the study from Markovits and Handley (2005) a similar pattern was observed, with the exception of ratings for MP, which were higher than for the other three inferences. Singman and Klauer (2011), using only a continuous response format, found a more pronounced effect of inference type. No effect of inference type can be expected in the framework of dual-process theories when people take a heuristic stance to the task. A heuristic stance could have been promoted in this experiment through the complexity of the task: In contrast to the two studies above, the relations employed here were arbitrary and each premise was provided with explicit probability information. This may have made it more difficult to explicitly both assess the entailment relations involved in the argument and integrate their probabilities. Thus, one could assess what effect results from simplifying the task, e. g. by providing probability information implicitly by using familiar conditional relations for which people readily build an idea of their probability. This would have the additional benefit that the validity and the soundness of the inferences would converge, ruling out the possibility that people's answers showed no effect of inference type because they were judging not their validity but their soundness, which in the arbitrary relations employed was set to be constant[2].

The main prediction of this study was that, provided people take an analytic stance to a problem, it would take longer to evaluate inferences when asked how probable it is that the conclusion is true, than when asked whether the conclusion follows from the premises, because integrating probabilities is an additional task to assessing entailment relations. In contrast, we predicted it to take less time to evaluate the truth of a statement when asked how probable it is that the statement is true than when asked whether the statement is true or false, since the latter would involve the additional task of setting a threshold and comparing it with the statements probability. We found that people where generally faster with continuous response format, and that when judging inferences, the entailment relations constituting the structure of the arguments had an effect for dichotomous but not for continuous response format. The results were in accordance with the hypotheses for statement evaluation, but not for inference evaluation. However, they suggest that when evaluating inferences, people did not take an analytic stance to the task when the response format was continuous. One way to promote an analytic stance could be through instructions introducing the task explicitly as one aimed at investigating how analytic reasoning differs from intuitive reasoning, making it important to engage in the former for the sake of the experiment. Such a manipulation was successful in eliciting a heuristic stance in a study from

[2] We thank Momme von Sydow for this helpful suggestion.

de Neys and Franssens (2009). One could then assess whether this would make a difference.

Acknowledgments

This work was supported by Grant KN 465/11-1 to MK from the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" (SPP 1516).

References

Adams, E. W. (1975). *The logic of conditionals: An application of probability to deductive logic.* Dordrecht, Holland: Reidel.

Baayem, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3*(2), 12-28.

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Luftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers, 28*(4), 584-589.

De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science, 7*(1), 28-38.

De Neys, W., & Franssens, S. (2009). Belief inhibition during thinking: Not always winning but at least taking part. *Cognition, 113,* 45-61.

Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review, 13*(3), 378-395.

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instructio- nal set on causal conditional inference. *The Quarterly Journal of Experimental Psychology, 63,* 892-909.

Geiger, S., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking,* (pp. 289-308). Oxford, UK: Oxford University Press.

Girotto, V., & Johnson-Laird, P. N. (2004). The probability of conditionals. *Psychologia, 47,* 207-225.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review, 109,* 646-678.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J. P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106,* 62-88.

Klauer, C. K., Beller, S., & Hütter, M. (2010). Conditional reasoning in context: A dual-source model of probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(2), 298-323.

Markovits, H., & Handley, S. (2005). Is inferential reasoning just probabilistic reasoning in disguise? *Memory & Cognition, 33*(7), 1315-1323.

Oaksford, M. & Chater, N. (2007). *Bayesian Rationality: The probabilistic approach to human reasoning.* Oxford, UK: Oxford University Press.

Oaksford, M., & Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 883-899.

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology, 53*(3), 238-283.

Oberauer, K., Geiger, S. M., Fischer, K., & Weidenfeld, A. (2007). Two meanings of "if"? Individual differences in the interpretation of conditionals. *The Quarterly Journal of Experimental Psychology, 60*(6), 790-819.

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

Singmann, H., & Klauer, K. C. (2011). Deductive and inductive conditional inferences: Two modes of reasoning. *Thinking & Reasoning, 17*(3), 247-281.

Soman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilo vich, D. Griffin, & D. Kahneman. Heuristics and Biases: The Psychology of Intuitive Judgment. Cambridge: Cam bridge University Press.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory and Cognition, 22,* 742-758. Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475-482.

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking and Reasoning, 1999, 5*(2), 97-113.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning, 11,* 239-278.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record, 58,* 475-482.

Wolf, A. & Knauff, M. (2008). The strategy behind belief revision: A matter of judging probability or the use of mental models? In B. C. Love, Kl McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 64-70). Austin, TX: Cognitive Science Society.