Uncertainty in Causal and Counterfactual Inference

Daniel Yarlett (dany@cogsci.ed.ac.uk)
Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW

Michael Ramscar (michael@dai.ed.ac.uk)
Division of Informatics, University of Edinburgh
2 Buccleuch Place, Edinburgh EH8 9LW

Abstract

We report 4 studies which show that there are systematic quantitative patterns in the way we reason with uncertainty during causal and counterfactual inference. Two specific type of uncertainty – uncertainty about facts and about causal relations – are explored, and used to model people's causal inferences (Studies 1-3). We then consider the relationship between causal and counterfactual reasoning, and propose that counterfactual inference can be regarded as a form of causal inference in which factual uncertainty is eradicated. On this basis we present evidence that there are also systematic quantitative patterns underlying counterfactual, as well as causal, inference (Study 4). We conclude by considering the consequences of these results for future research into causal inference.

Introduction

The ability to make causal inferences is of central importance to cognitive agents wishing to control or predict events in the world. However, many of our beliefs are held with less than perfect certainty. Given this, it is natural to enquire about the way in which uncertainty affects the process of reasoning about the world. In this paper we examine the way in which two types of uncertainty – uncertainty about facts and uncertainty about causal relations – are assimilated during the process of causal and counterfactual inference. Studies 1-3 reveal that there are systematic quantitative patterns in our treatment of uncertainty in causal inference, suggesting that our understanding of causality is not inherently deterministic as has recently been proposed (Goldvarg & Johnson-Laird, 2001). We then consider the relationship between causal and counterfactual inference, and show that modified forms of the models which can be used to predict causal inference can be used to predict counterfactual inference (Study 4), a result consistent with theories which treat counterfactuals as supervenient on causal knowledge (Pearl, 2000; Yarlett & Ramscar, in press; Jackson, 1977). We conclude by considering the consequences of these results for future research into causal and counterfactual inference.

Causal Inference and Uncertainty

We make a causal inference when we acquire some new evidence about a cause, and on this basis update our beliefs about effects related to that cause. For example, imagine you meet Tom at a party. During your brief conversation, he says and does certain things that make you think he is an Army Officer, although you're not completely certain about this. As a result of this suspicion you might now think it more likely that Tom is able to fire a pistol and abseil, compared to when you first met him. Your beliefs about Tom have changed as a result of causal inference.

When it comes to making causal inferences, two types of uncertainty are especially important: *factual uncertainty* and *causal uncertainty*. Factual uncertainty arises simply because our experience of the world is in many cases insufficient to allow us to be completely certain about our beliefs. For example, Tom's extensive knowledge of firearms and military strategy, as displayed in your conversation, might make you suspect that he is in the Army. But you are nevertheless aware that you could be wrong about this. Therefore there is some factual uncertainty in your belief that Tom is in the Army.

The second type of uncertainty relevant to making inferences about the world is *causal uncertainty*. This arises because although there are systematic regularities in the world, these rarely obtain without exception. For example, we all agree that clouds cause rain, even though rain does not invariably fall when it is cloudy; and we would probably also concur that smoking causes cancer, although we know that not all smokers contract cancer. Causal uncertainty, then, arises because of our awareness that although events of type A may tend to produce events of type B, it is not the case that As are *always* or *invariably* followed by Bs.

Although it seems intuitively plausible that both factual and causal uncertainty should play a role in determining our causal inferences, to our knowledge very little empirical work has explored this issue. Some previous work has found an effect of factual uncertainty in both deductive (Stevenson & Over, 1995; Byrne, 1989) and causal (Cummins *et al.*, 1991) settings, but none of these studies examined the systematic effects of factual uncertainty from a quantitative perspective. And although it seems reasonable to assume that causal uncertainty plays a role in causal inference and reasoning – and indeed, many recently proposed theories (e.g., Cheng, 1997; Pearl, 2000) and models (Rehder, 1999; Yarlett & Ramscar, in press) concerned with causal reasoning successfully make this assumption – it is by no means uncontentious. Goldvarg & Johnson-Laird (2001) have recently argued that the meaning of causal statements is inherently deterministic, and more generally, theories of reasoning which invoke mental models do not easily permit the accommodation of less than certain inferences (but see Johnson-Laird, 1994, and Stevenson & Over, 1995). The present series of studies therefore set out to investigate whether factual and causal uncertainty play a role in the process of causal inference and, if so, whether they do so in a systematic fashion.

Study 1

Study 1 was designed in order to get ratings about the causal uncertainty attaching to a specific set of cause-effect pairs, in order to explore the structure of the information with which people relate causes and effects, and also to investigate the information that might be used in causal inference. People were asked to rate the causal uncertainty attaching to a range of cause-effect pairs on a range of scales which measured: (i) how strongly the cause causes the effect; (ii) how strongly the effect depends on the cause; (iii) the conditional probability of the effect given the presence of the cause; and (iv) the conditional probability of the effect given the absence of the cause. In addition to the ratings collected, the following ratings were derived from the conditional probability ratings:

$$\Delta P \text{ Contingency} = P(e|c) - P(e|{\sim}c)$$

$$\text{Power PC} = \frac{P(e \mid c) - P(e \mid {\sim}\, c)}{1 - P(e \mid {\sim}\, c)}$$

These quantities have variously been proposed as measures of the strength of a cause (e.g. Cheng & Novick, 1992; Cheng, 1997).

Materials and Design. The materials used described 10 different cause-effect pairs. They were selected in order to cover a wide variety of domains, and included the following pairs: smoking and cancer; cars and pollution; stress and insomnia; sunbathing and suntanning; weight-training and muscle-growth; cholesterol and heart attacks.

Subjects were asked directly about the strength of relation that they thought held between the pairs in question. For example, for the smoking-cancer pair, the following questions were used: (i) "How strongly do you think smoking causes cancer?"; (ii) "How strongly do you think whether someone gets cancer depends on whether they smoke?"; (iii) "How likely do you think someone would be to get cancer given that they smoke?"; and (iv) "How likely do you think someone would be to get cancer given that they do not smoke?" All ratings were collected on a 0-100 scale. For the causal ratings the scale was anchored by 'does not cause at all' and 'always causes'; for the dependency ratings 'does not depend at all' and 'perfectly depends'; and for the subjective probability ratings 'completely unlikely' and 'completely certain'.

Three groups were asked to rate the causal, dependency, and conditional probability ratings. A within-subjects design was not used because of concerns that this would artificially homogenise what might in reality be different ratings (e.g. being asked to rate causal, dependency and conditional probability ratings consecutively might encourage subjects to simply return similar ratings on all scales).

Participants. 49 students from the University of Edinburgh participated voluntarily.

Results. A factor analysis (principal-components analysis with rotated axes) was conducted on the ratings in order to examine their structure. Only the first two rotated factors had eigenvalues greater than 1, and these together accounted for 95.16% of the variance in the data. The factor loadings are shown in Table 1.

	Factor 1 (Causal Power)	Factor 2 (Baserate)
Causal	0.969	-0.042
P(e|c)	0.927	-0.326
Power PC	0.873	-0.481
Dependency	0.699	-0.600
ΔP	0.664	-0.738
P(e|~c)	-0.106	0.982

Table 1: Factor loadings from Study 1.

Discussion

The two factors extracted in the factor analysis successfully explained a large proportion of the variance in the causal uncertainty ratings collected in Study 1. Moreover, the extracted factors are readily interpretable because the causal ratings load very highly on the first factor (0.969) and negligibly on the second factor (-0.042), while the P(e|~c) ratings load very highly on the second factor (0.982) and negligibly on the first factor (-0.106). Study 1 therefore suggests that two factors are especially important in accounting for our representation of causal uncertainty: the causal strength with which a cause produces its effect ('Causal Power'), and the base rate of the effect in the absence of the cause ('Baserate'). This suggests that models of

Model	Definition
Probabilistic	$P(e\|c)P(c) + P(e\|{\sim}c)P({\sim}c)$
Linear	$P(e\|{\sim}c) + causes(c,e)P(c)$
Noisy-OR	$1 - [1 - P(e\|{\sim}c)][1 - causes(c,e)P(c)]$
Causal	$causes(c,e)P(c)$
Dependency	$depends(e,c)P(c)$

Table 2: The models of causal inference.

causal inference should incorporate these two parameters. This proposal was investigated in Study 2.

Study 2

Study 2 examined the degree to which causal inference can be modelled using information about factual and causal uncertainty. The ratings from Study 1 were used in order to provide information about the degree of causal uncertainty attaching to the 10 causal pairs, while new data was acquired concerning their factual uncertainty. Short scenarios centring around each of the 10 causal pairs were designed, in which it was deliberately made unclear whether the cause was present or absent. These were used to induce factual uncertainty in participants in the study. For example, the scenario for the smoking-cancer causal pair ran as follows:

> "Imagine you're introduced to Bill, a friend of a friend, one day. You ask Bill for a lighter but he doesn't carry one. However, it does look a little as though he might have tobacco stains under his nails."

After reading each description, participants were requested to rate their factual uncertainty, on a 0-100 scale, by being asked how likely they thought it was that the cause was present given what they had read (i.e., in this case how likely they thought it was that Bill was a smoker). They were then asked to make a causal inference by judging, given their confidence that Bill may or may not be a smoker, how likely they thought he would be to contract cancer at some point in his life. The information collected about factual and causal uncertainty was then used to parameterise various models of causal inference, in order to see if the inferences participants made could be predicted.

Models of Causal Inference

The models of causal inference investigated are listed in Table 2. The probabilistic model defines the normative method of inferring the probability of an effect given information about a related cause. The linear model, in contrast, states that one's belief in an effect is the combination of a base rate of belief – the belief that the effect is present in the absence of the cause – plus the extra support that the cause provides for belief in the effect, which is defined as the product of one's belief that the cause is present and the degree to which the cause and effect are causally related. The noisy-OR model (Pearl, 1988) treats causes as mechanisms that operate independently and additively to produce a common effect. The probability of an effect in this framework is thus given as the probability that not all the causes fail to generate the effect.[1] Finally, the causal model predicts that people's belief in the cause is a product of the degree to which the cause and effect are causally related, and the degree to which the cause is believed to be present. And the dependency model is similar to the causal model, except that it measures causal uncertainty using dependency, instead of causal strength, ratings.

Materials and Design. The cause-effect pairs and connection ratings from Study 1 were used. In addition, scenarios for each causal pair were designed in order to embed the causal relation in a specific context, and deliberately induce factual uncertainty as to whether the cause in question was present or not.

A within-subjects design was deliberately eschewed in Study 2 because of concerns that it could artificially bring people's causal inferences in line with the predictions of the probabilistic model. Many people are familiar with basic probability theory, and our concern was that being asked to rate the conditional probability of the effect given the cause before making their causal inference (as a within-subjects design would have required), could force people to reason about the effect arithmetically, in opposition to their natural style of reasoning. Accordingly, a between-subjects design was adopted, in which the causal uncertainty ratings used were those collected in Study 1, while the factual uncertainty ratings and the causal inferences themselves, were collected in the present study.

Participants. Participants were 21 students at the University of Edinburgh. All participants were volunteers, and no reward was offered for participation.

Results. The performance of the models of causal inference is shown in Figure 1. The probabilistic ($r = 0.665$, $\underline{p} < 0.05$, one-tailed), linear ($r = 0.621$, $\underline{p} < 0.05$, one-tailed), and noisy-OR ($r = 0.711$, $\underline{p} < 0.05$, one-tailed) models were all significant predictors of people's causal inferences. The causal ($r = 0.495$, $\underline{p} > 0.05$, one-tailed) and dependency ($r = 0.268$, $\underline{p} > 0.05$, one-tailed) models, however, failed to significantly predict people's inferences. A further analysis was also

[1] Interestingly, the linear and the Noisy-OR models of causal inference find their counterparts in the ΔP and Power PC theories of causal induction respectively (they can be derived as the maximum likelihood estimates of causal strength parameters in causal graphs appropriately parameterised; see Glymour, 1998; Tenenbaum & Griffiths, 2000).

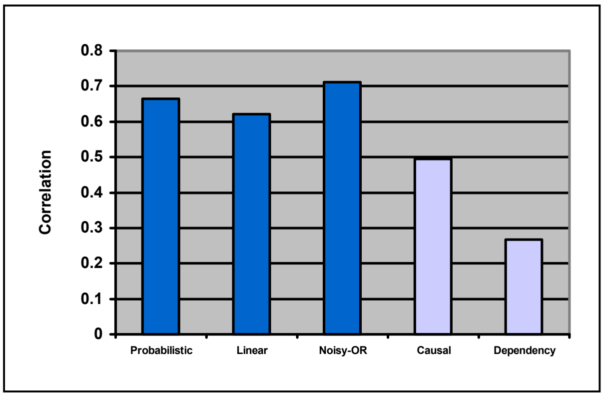

Figure 1: Results of Study 2.

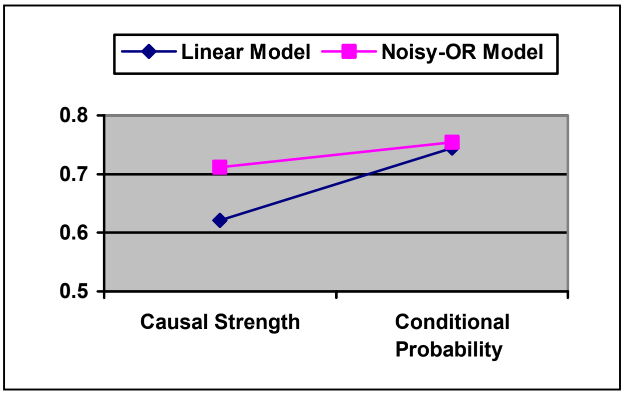

Figure 2: Effect of changing parameterisation of Linear and Noisy-OR models.

conducted on the linear and noisy-OR models, because they could also be parameterised with conditional probability instead of causal strength ratings. In both cases, parameterising the models with the conditional probability of the effect given the presence of the cause, instead of the causal strength ratings, served to increase their empirical performance (see Figure 2).

Discussion

The fact that the linear, noisy-OR and probabilistic models were significantly correlated with the strength of people's causal inferences suggests that information about factual and causal uncertainty plays an important role in the inference process, and also that there seem to be domain-general quantitative patterns in the way we reason from cause to effect. However, the factual and causal uncertainty ratings and inferences predicted in Study 2 were between-subjects aggregates. It is therefore possible that the success of the proposed models is merely an artefact of the experimental design, and that the models would prove unable to predict causal inferences on a within-subjects basis. Study 3 investigated this issue, while also allowing us to examine how much of the residual error in the causal models could be attributed to idiosyncratic use of the rating scales.[2]

Study 3

Study 3 used a within-subjects design in which people estimated the factual and causal uncertainty attaching to each of the 10 cause-effect pairs, and then made a causal inference about the effect. Because the causal and dependency models failed to significantly predict people's inferences they were dropped from consideration. Instead we focused on the performance of just the probabilistic, linear and noisy-OR models.

Materials and Design. The causal pairs and materials as used in Studies 1 and 2 were again used in this study. Each subject saw all 10 scenarios, in one of two reverse-orderings. The linear and noisy-OR models were parameterised using only conditional probability, and not causal strength ratings, because of the better performance of this form of the models in Study 2.

Participants. Participants were 15 students enrolled at the Division of Informatics, University of Edinburgh. All participants were volunteers, and no reward was offered for participation.

Results. The performance of the causal models is shown in Figure 3. Both the linear ($t = 2.280$, $df = 14$, $p = 0.038$, two-tailed) and the noisy-OR model ($t = 2.379$, $df = 14$, $p = 0.032$, two-tailed) performed significantly better than the probabilistic model, although there was no significant difference between the linear and noisy-OR model ($t = 1.302$, $df = 14$, $p = 0.214$, two-tailed). The degree of variance explained in the inference process by just taking into account either the amount of factual uncertainty, in the form of the $p(c)$ ratings, or the amount of causal uncertainty, in the form of the $p(e|c)$ ratings, is also shown in Figure 3 for comparison. These two models performed significantly worse than all the other models.

To confirm that both the factual and causal uncertainty parameters added to the models' predictive validity the performance of the linear and noisy-OR models was compared to modified versions of them in which (i) factual uncertainty was ignored; and (ii) causal uncertainty was ignored. The linear model performed significantly better than its counterpart which ignored factual uncertainty ($t = 2.358$, $df = 14$, $p = 0.017$, one-tailed), and marginally better than its counterpart which ignored causal uncertainty ($t = 1.546$, $df = 14$, $p = 0.072$, one-tailed). The noisy-OR model performed significantly better than both its modified

[2] Because Study 2 had shown that the probabilistic model predicted causal inference with some level of success in a context in which patterns of causal inference consistent with the predictions of the probabilistic model could not have been artificially induced, the use of a within-subjects design was now appropriate.

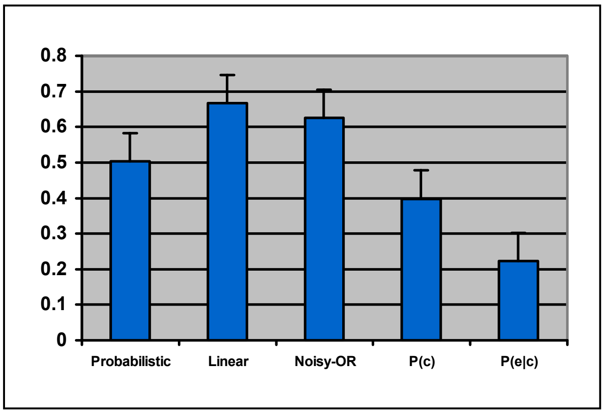

Figure 3: Results from Study 3.

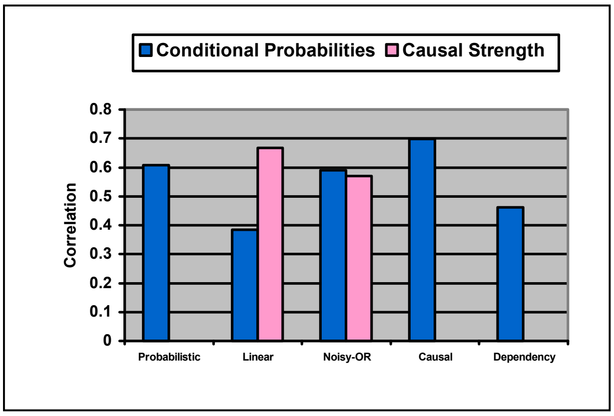

Figure 4: Results from Study 4.

versions which ignored factual ($t = 4.199$, $df = 14$, $p < 0.001$, one-tailed) and causal ($t = 2.049$, $df = 14$, $p = 0.030$, one-tailed) uncertainty.

Discussion

The results of Study 3 show that quantitative models – particularly the linear and noisy-OR model – can successfully predict people's causal inferences with some degree of success. Moreover, the results of Study 3 also show that removing information about either factual or causal uncertainty from these models significantly decreases their performance, thus showing that these factors do seem to play an important role in causal inference.

Causes and Counterfactuals

The studies reported so far examined the role of uncertainty in causal reasoning. However, there is an intimate connection between causal and counterfactual reasoning (c.f. Lewis, 1973b; Jackson, 1977; Pearl, 2000; Yarlett & Ramscar, 2001). In the light of this it is interesting to examine whether the findings concerning causality in Studies 1-3 can also be applied to counterfactual reasoning.

The proposal we examined is that, at least in the present context, counterfactual reasoning can be treated as a form of causal reasoning in which residual factual uncertainty is eliminated (for treatments of counterfactual reasoning in more complex systems see Yarlett & Ramscar, in press, and Pearl, 2000). For example, imagine that you are fairly sure that Bill is not a smoker, but that I ask you how likely you think he would be to contract cancer if (counterfactually) he were a smoker. Even though there may be some factual uncertainty in your belief that Bill is not *actually* a smoker, there should be no factual uncertainty attaching to the counterfactual scenario because the counterfactual asks you to assume, unequivocally, that he *is* a smoker. Study 4 investigated this proposal.

Study 4

Study 4 investigated whether quantitative patterns could be found underlying counterfactual, as well as causal, inference. The scenarios used in Studies 2 and 3 were altered so that instead of engendering uncertainty they were perfectly unambiguous that the cause in question was absent. Then, instead of being asked to make a straightforward causal inference, subjects were asked to consider how strongly they would believe in the effect *if the cause were present.*

Materials and Design. The materials used were adapted forms of the scenarios used in Studies 2-3. Here is the smoking scenario, with the added information shown in italics:

> "Imagine you're introduced to Bill, a friend of a friend, one day. You ask Bill for a lighter but he doesn't carry one. However, it does look a little as though he might have tobacco stains under his nails. *It later turns out that Bill is not a smoker; in fact he's never even smoked a cigarette in his life.*"

Subjects were then asked to rate "But if Bill were a smoker, how likely do you think he would be to get cancer at some point in his life?". Data was collected using a between-subjects design, as used in Study 2.

Participants. Participants were 23 students at the University of Edinburgh.

Results. The results of Study 4 are shown in Figure 4. The causal model ($r = 0.699$, $df = 8$, $p < 0.05$, one-tailed), linear model parameterised with causal strength ratings ($r = 0.667$, $df = 8$, $p < 0.05$, one-tailed), and noisy-OR model parameterised with either conditional probabilities ($r = 0.589$, $df = 8$, $p < 0.05$, one-tailed) or causal strengths ($r = 0.571$, $df = 8$, $p < 0.05$, one-tailed), significantly predicted people's counterfactual inferences.

Discussion

The results of this study show that modified forms of the models used to predict causal inferences can also be employed in the prediction of counterfactual inferences, and also that counterfactual inference can be profitably regarded as a special case of causal inference in which factual uncertainty has been eradicated. This result is both consistent with theories which hold that counterfactuals supervene on causal relations (e.g., Jackson, 1977; Pearl, 2000; Yarlett & Ramscar, in press), and at tension with theories that treat counterfactual judgements as propositions assigned binary truth-values (e.g., Byrne, 1997; Byrne & Tasso, 1999; Lewis, 1973). However, given the success of multiple models at capturing the quantitative patterns in counterfactual inference exhibited in Study 4, clearly further work is required to tease the models apart, and determine whether patterns in both causal and counterfactual inference can be successfully captured by the same models.

General Discussion

The 4 studies reported here suggest that both factual and causal uncertainty play an important role in determining causal and counterfactual inference, and furthermore that counterfactual inference can profitably be regarded as a form of causal inference in which factual uncertainty is eradicated. However, one potential cause for concern is the often considerable amount of variance left unexplained by the sort of quantitative models described in this paper. Clearly more work needs to be done before the role of such models in describing causal and counterfactual inference is fully understood. In particular, in future work we intend to examine whether alternative ways of measuring causal and factual uncertainty can increase the explanatory power of the quantitative models, and also whether additional factors can be imported into the models to improve their empirical fit (e.g. how many alternative or preventative causes exist for a specific cause effect pair being reasoned about; see Cummins *et al.*, 1991).

Acknowledgements

The authors would like to thank attendees of the workshop on *Causal Learning and Inference in Humans and Machines* at NIPS*2001.

References

Byrne R.M.J. (1989). Suppressing Valid Inferences with Conditionals, *Cognition*, **31**, 61-83.

Byrne R.M.J (1997). Cognitive Processes in Counterfactual Thinking About What Might Have Been, *Psychology of Learning and Motivation*, **37**, 105-154.

Byrne R.M.J. and Tasso A. (1999). Deductive Reasoning with Factual, Possible, and Counterfactual Conditionals, *Memory & Cognition*, **27**(4), 726-740.

Cheng P.W. (1997). From Covariation to Causation: A Causal Power Theory, *Psychological Review*, **104**(2), 367-405.

Cheng P.W. and Novick L.R. (1992). Covariation in Natural Induction, *Psychological Review*, **99**(2), 365-382.

Cummins D.D., Lubart T., Alksnis O. and Rist R. (1991). Conditional Reasoning and Causation, *Memory & Cognition*, **19**(3), 274-282.

Glymour C. (1998). Learning Causes: Psychological Explanations of Causal Explanation, *Minds and Machines*, **8**, 39-60.

Goldvarg E., and Johnson-Laird P.N. (2001). Naïve Causality: A Mental Model Theory of Causal Meaning and Reasoning, *Cognitive Science*, **25**, 565-610.

Hadjichristidis C., Stevenson R.J., Over D.E., Sloman S.A., Evans J.St.B.T., Feeney A. (2001). On the Evaluation of *If p then q* Conditionals, *Proceedings of the Twenty-third Annual Conference of the Cognitive Science Society* (pp. 381-386). Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Jackson F. (1977). A Causal Theory of Counterfactuals, *Australasian Journal of Philosophy*, **55**, 3-21.

Johnson-Laird P.N. (1994). Mental Models and Probabilistic Thinking, *Cognition*, **50**, 189-209.

Lewis D.K. (1973a). *Counterfactuals*, Blackwell, Oxford, UK.

Lewis D.K. (1973b). Causation, *Journal of Philosophy*, **70**, 556-567.

Rehder B. (1999). A Causal Model Theory of Categorization, *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 595-600). Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Pearl J. (1988). Probabilistic Inference for Intelligent Systems, Morgan Kaufmann, San Mateo, California.

Pearl J. (2000). Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, UK.

Stevenson R.J. and Over D.E. (1995). Deduction from Uncertain Premises, *The Quarterly Journal of Experimental Psychology*, **48A**(3), 613-643.

Tenenbaum J.B. and Griffiths T.L. (2000). Structure Learning in Human Causal Induction, *Advances in Neural Information Processing Systems*, **13**, MIT Press, Cambridge, Massachusetts.

Yarlett D.G. and Ramscar M.J.A. (in press). A Quantitative Model of Counterfactual Reasoning, *Advances in Neural Information Processing Systems*, **14**, MIT Press, Cambridge, Massachusetts.