The Relationship between Causal and Counterfactual Reasoning

William Jimenez-Leal (W.Jimenez@warwick.ac.uk)
Department of Psychology, University of Warwick, Coventry. CV5 7 AL, England.

Nick Chater (N.Chater@ucl.ac.uk)
Department of Psychology, University College London, Gower Street, London. WC1E 6BT, England.

Abstract

In this paper it is claimed that counterfactual reasoning in contextualized situations depends on and reflects causal contingencies, which are actualized depending on the task demand. The experiments presented manipulated some elements of the pragmatics of a task to show cases where dissociation between causal and counterfactual reasoning does or does not occur. Based on this evidence, it is claimed that Judgement Dissociation Theory (Mandel, 2003b) does not adequately characterize the link between causal and counterfactual reasoning. Results are discussed in terms of the pragmatics associated with causal queries.

Keywords: Counterfactual reasoning; causal reasoning; Norm Theory

Introduction

The objective of the following experiments is to identify some of the factors that determine the extent of the relationship between causal and counterfactual reasoning. The nature of the link between these types of reasoning is still not clear, despite a renewed interest in the subject. (Byrne, 2002; Spellman, Kincannon, & Stose, 2005). An influential view (Mandel, 2003b) claims that causal and counterfactual reasoning are actually independent and dissociated.

We will first present a summary of JDT and some alternatives to it. Then we will examine the problems of JDT and formulate an alternative explanation for its supporting evidence. The experiments reported are based on that reinterpretation.

Judgement Dissociation Theory (JDT)

Mandel (2003b) puts forward JDT as a way to conceptualize the apparent independence between the process of causal selection and counterfactual reasoning. The operation of JDT operation is based on two principles, actuality and substitution. According to the actuality principle, causal selection is based on the existence of sufficient antecedents to the actual outcome of the case under consideration. According to the substitution principle, counterfactual and covariational reasoning depend on the generation of *ad hoc* categories (or norms) that systematically focus on elements that can undo an outcome, called 'preventors'. The two principles operate independently, are based on different information (Mandel & Lehman, 1996), have different targets, and do not influence each other (Mandel, 2003a). Whereas causal attributions are determined by the identification of the process that actually produces a specific outcome (mechanism beliefs), counterfactuals are governed by the creation of *ad hoc* categories depending on the short term objectives of the reasoner.

Mandel (2003b) proposes JDT in contrast with the approaches he dubs Spellman Probability Updating Account (SPU) and Causal Simulation Approach (CSA). The first represents a covariational model and the second, a counterfactual model of causal selection. Mandel (2003b) presents evidence against these approaches.

SPU is proposed by Spellman (1997) to explain 'token causation' in an analogous way to the covariational approach to type causality (Cheng, 1997). Spellman (1997) proposes that causality judgments are a function of the probability increase of the outcome above its prior probability: $C=p(O_{After})-p(O_{Before})$. The probability estimates are derived from pre-existing knowledge (including covariation and causal mechanisms (Spellman et al., 2005) and counterfactual reasoning. More specifically, counterfactual judgements influence the estimates of baseline probabilities, although Spellman does not specify how.

SPU also claims that causal and counterfactual judgments are based on the same information (prior causal knowledge), represented by the prior probability of the outcome, and accordingly, an event undone by a counterfactual should be considered an effective cause if it also represents an increase over the base rate probability.

CSA is a label given by Mandel to analyses that rely on counterfactual simulations to identify causes. According to CSA, people identify the causes of an event by performing a mental simulation of the negation of the candidate cause. If as a result of the simulation the effect is suspended, then the first event will be selected as causally effective (Wells & Gavanski, 1989; Wells, Taylor, & Turtle, 1987). This analysis is inspired by the philosophical counterfactual analysis of causation (See Collins, Hall, & Paul, 2004) and based more directly on Kahneman and Tversky's simulation heuristic (Kahneman, et al. 1982) and Norm Theory (Kahneman & Miller, 1986). Some consequences of this perspective include the idea that causation is a relation of necessity, not sufficiency; and that understanding counterfactual statements is equivalent to understanding causal statements.

JDT uses cases of pre-emption to analyse and criticize both SPU and CSA accounts, echoing the discussion in philosophy. An example is the case of Suzy and Billy. Billy throws his rock first but Suzy's rock hits the bottle first, breaking it. In this case, SPU would incorrectly single out Billy's throw as the cause, since the probability of the bottle breaking was higher after Billy threw the rock than before. CSA also fails to attribute causality, since both

counterfactual simulations of the absence of candidate causes fail to undo the effect. Consider statements (1) and (2), as representing the corresponding simulation.

1. If Suzy hadn't thrown the rock, the bottle wouldn't have broken.
2. If Billy hadn't thrown the rock, the bottle wouldn't have broken.

In both cases the statements are false, indicating that eliminating the candidate cause in each case does not eliminate the effect.

In contrast, JDT's actuality principle allows reasoners to identify the specific event that brings about an effect because people acknowledge sufficiency as the hallmark of causality, an element not represented in either CSA or SPU. The substitution principle predicts that people will focus on preventors, elements that are also out of the scope of the rival theories.

The nature of the dissociation

Some problems can be identified in the formulation of JDT that cast doubt on the extent of the dissociation between causal and counterfactual reasoning. For example, the key concepts of 'sufficiency' and 'preventor' are underspecified. There is no broad agreement on how to define them, and more importantly, it has been shown that the concepts of formal sufficiency and necessity (on which JDT is based) do not match people's understanding of them (Verschueren et al, 2006). Some researchers have convincingly argued in favour of a contextual definition of sufficiency and necessity (See Hart & Honoré, 1985; Hilton & Erb, 1996; Hilton et al, 1990), and although Mandel's account of causal selection explicitly acknowledges it as a conversational process, he does not specify any element of the conversational process that might influence causal selection. Thus, the conversational processes inherent to causal selection, as well as the characterization of sufficiency and necessity under the circumstances, are left unexplained.

On the other hand, JDT's concept of counterfactual reasoning is based on an illicit generalization of the idea of the simulation heuristic (Kahneman et al., 1982). Within Norm Theory (Kahneman & Miller, 1986), the simulation heuristic was proposed as a post factor reaction intending to undo a negative outcome determined by abnormal events (Kahneman & Miller, 1986). In that sense, counterfactual thinking aiming to restore a norm[1] should be motivationally relevant.

Counterfactual thinking whose objective is to test a causal structure is not necessarily constrained by this element, although it is clear that both may concur in some situations. In some cases counterfactual backward processing requires counterfactual forward thoughts in order to test the contingencies that can undo a target, but these counterfactual thoughts are implicitly revealed (Kahneman, 1995). In other words, eliminating an outcome is not the same as explaining it causally, and not all counterfactual simulations aim to identify causes. The element common to all forms of counterfactual thinking involves a mental simulation that reveals, often implicitly, causal knowledge in the form of the rules that govern the simulation (Kahneman, 1995). This causal texture is implicitly displayed and only explicitly used depending on the demands of the task.

Even when counterfactual reasoning is required to answer a causal question, the interpretation of the question plays a key role in determining the process required. By taking the operation of the simulation heuristic out of context, JDT equates its operation with all counterfactual reasoning and fails to acknowledge that counterfactual contingencies are actualized by the demands of the task.

Finally, by virtue of the actuality principle, JDT privileges causal mechanism information in the process of causal selection, in contrast with probabilistic information. JDT claims that probabilistic information is not relevant for particular cases, only for 'type' causation. However, causal selection can rely on either of these sources, and what is more, in many cases they are equivalent (Cheng & Glymour, 1998). Where JDT aims to describe different questions (how vs. what), it describes different sources of information. The difference lies in the task demand, not in the information itself.

When causal and counterfactual thoughts coincide

There is evidence that points to integration, rather than to dissociation, of causal and counterfactual reasoning. The appropriate modelling of the causal structure might help to understand how causal and counterfactual reasoning relate, and at the same time provide a normative framework for studying counterfactuals (Sloman & Lagnado, 2005).

According to the causal modelling framework (Glymour & Cooper, 1999; Pearl, 2000), the causal structure of a situation constrains the kind of counterfactual inferences that are allowed. Sloman and Lagnado (2005) have used it to explore the issue of counterfactual reasoning in deterministic causal systems. Their main finding is that when reasoning about the consequences of a counterfactual supposition, people do not alter their beliefs about the state of the normal causes of the outcome. That is, the causal structure against which counterfactuals are judged is kept stable. Moreover, Sloman and Lagnado found that people correctly identify the outcome of imagined interventions based on counterfactual assumptions. Their conclusion is that causal inference follows the logic of intervention; causal inference is determined by counterfactually altering the values of a variable that is part of a causal network.

How is it possible then to reconcile these findings with JDT? Spellman et al (2005) proposed that the dissociation is an order effect. In cases where the same participants have to complete causal and counterfactual tasks, when subjects are asked first to generate counterfactuals, that information becomes available for causal evaluation, making it more probable to affect performance in the causal task. However, when the causal task is performed first, this does not affect mutations performed afterwards. However, contrary to

[1] A norm is to be defined according to a framework of normality based on elements like closeness, values and moral evaluations.

Spellman et al. (2005), Mandel (2003a; 2003b) did not find any order effects. Spellman et al's explanation is tested in these experiments.

An alternative explanation is that the mismatch observed by JDT occurs simply because causal and counterfactual queries are not usually specified in the same way. In fact, Mandel's (2003b) counterfactual probes always refer to undoing the 'outcome of a situation' whereas causal questions refer to a particular event (e.g. 'glass bottle breaking yesterday') (Mandel, 2003a, p. 423). Similarly, the instructions for the ratings requested were not consistent. Second, the mismatch in the description of the events in the tasks also leads to obscuring the underlying causal structure of the situation. Once the 'causal model' is clear, it is feasible that counterfactual and causal tasks can have the same targets. Causal queries convey cues that somehow specify the content of the causal answers intended to be received. A question about the cause of a theoretical outcome has more room for interpretation than a question about the cause of a 'glass bottle breaking yesterday'.

In summary, ambiguous elements that involve a certain degree of pragmatic interpretation can be responsible for some of the cases of dissociation between causal and counterfactual reasoning. The experiments reported below manipulate these elements, the probe and the specificity of the description, using Mandel's (2003b) original materials and methods to contrast the results.

In the scenario presented, a criminal falls prey to two assassination attempts. The first assassin puts poison in his drink, which should take one hour to have any effect. However, before the poison has killed him, the second assassin runs the criminal off the road. The criminal dies because of the explosion of the car.

Mandel's participants consistently chose and rated the main character's involvement in crime as the best way of undoing the 'outcome of the story' (counterfactual target) and consistently chose and rated the car crash as the 'cause of his premature death' (causal target). In the following experiment the causal and counterfactual questions were matched at three levels of specificity, and it is expected that target selection and rating will be matched at the most specific level and dissociated at the more general level of description (Table 1). JDT does not specify the influence of this factor and it is assumed that it predicts dissociation regardless of the specificity of the question posed.

Table 1. Sample of the original and alternative phrasing of the questions in Experiment 1.

	Causal	Counterfactual
Original phrasing	List factors considered "as causes of Mr. X's death."	"ways of undoing the story so the outcome would be different"
Low specificity	List factors considered "as causes of the outcome o the story"	"ways of undoing the outcome of the story"
Medium specificity	List factors considered "as causes of Mr. X's death."	"ways of undoing Mr X's death"
High specificity	List factors considered "as causes of Mr. X's death by fatal burns"	"ways of undoing Mr. X's death by fatal burns"

Experiment 1

Dissociation between causal and counterfactual reasoning is predicted to depend on the level of specificity of the description of target events in each task. Results of the experiment are interpreted by contrasting JDT with rival hypotheses. The presence of order effects, as suggested by Spellman et al. (2005), is also investigated.

Method

72 undergraduate students of different backgrounds from the University of Warwick took part in the experiment in exchange for payment. Participants were tested individually in a cubicle with a computer-based experiment. The complete display followed the structure of Mandel's (2003b) experiments, and included a causal, a counterfactual and a probability task. Participants were first presented with the general instructions, where they were told that they would have the opportunity to read a vignette and then asked questions about it.

Once the participants had read the scenario, they would proceed to complete the tasks. The order of the tasks was randomly counterbalanced. The causal and the counterfactuals tasks consisted of option listing and rating the answers participants wrote. The counterfactual task exhibited the same structure, with participants first asked to propose four ways in which the *event* would have been different, and then invited to rate each one.

The description of the 'event' varied between the three levels mentioned and for each level the description was matched across tasks. Participants were randomly assigned to one of the three versions defined by the specificity level.

In the probability task, participants were asked to estimate probabilities for the outcome, given four conditions defined by the presence/absence of the actions of the assassins: None of them occurring; Poisoning but no crash occurring; crash but no poisoning occurring; both occurring.

Results

Answers were coded by the author and an independent rater, who used the predefined categories of interest: 'crime life', 'poison', 'crash' and 'poison and crash'.

Proportions of participants and importance ratings

Table 2 presents a summary of the percentage of participants and the importance ratings for each target across specificity level for both types of judgements. Contrary to JDT, *crime life* was not the preferred modal response for the counterfactual task for any of the levels. In fact, in the high specificity level very few people chose it as a way of undoing the event. The proportion of people who chose *crime life* at this level, as either a cause or a way of undoing the target event, is significantly lower than the proportion at the low and medium levels (causal [$X^2(2, n = 28) = 6.2$, $p < .05$]; counterfactual [$X^2(2, n = 41) = 5.9$, $p < .05$])

Table 2 Percentage of participants and average ratings as a function of the task and specificity level.

		Judgement Type			
		Counterfactual		Causal	
Target	Level	%	M	%	M
Crime life	Low	54%	7.4	63%	6.7
	Medium	50%	6.7	75%	6.8
	High	13%	5.0	33%	8.3
Poison	Low	58%	4.7	42%	6.7
	Medium	70%	4.0	70%	5.9
	High	83%	5.6	75%	4.8
Crash	Low	63%	5.6	70%	7.0
	Medium	75%	3.1	70%	7.4
	High	83%	4.1	75%	7.4
Crash and Poison	Low	17%	7.9	4%	5.0
	Medium	20%	5.7	8%	5.5
	High	33%	6.9	8%	7.7

Crash and poison follows a similar pattern for both the causal and the counterfactual tasks. Similar proportions of people considered these targets to be the cause, independent of the level [$X^2(2, n=5) = 0.4$, $p = .8$].

Overall, it can be seen that proportions of people choosing a target are fairly similar across tasks, and that the higher number of people for both tasks is concentrated around the *poison* and *crash* targets. There are no significant differences between the number of participants who chose *poison* or *crash* as cause of the events across levels (poison [$X^2(2, n = 45) = 2.4$, $p = .29$] ; crash [$X^2(2, n = 52) = .05$, $p = .97$]). This seems to indicate that people did not consider the effect of these factors to be independent, and in any case *crash* alone was not considered to be sufficient to produce the 'event'. In the counterfactual task the same pattern emerges. (poison [$X^2(2, n = 51) = 1.1$, $p = .59$]; crash [$X^2(2, n = 53) = 0.6$, $p = .71$]).

Although the percentage of participants does not vary much within each level of specificity for the counterfactual task, the causal ratings do. A mixed ANOVA was conducted on the importance ratings (2 (judgement type) x 4 (target) x 3 (specificity level)). There is an interaction between target and judgement type, [$F(3, 67) = 8.60$, MSE = 78.20, $p < .01$] and although there is no main effect of the specificity level [$F(2, 69) = .26$, MSE = 2.7, $p = .76$], there is interaction between the specificity level and the target [$F(6, 62) = 2.15$, MSE = 29.92, $p < .05$]. That is, the ratings assigned to the targets varied as a function of the level of description. Post hoc comparisons performed on the counterfactual ratings showed that *crime life* is considered more effective in undoing the event in the low and medium levels, that is, the more ambiguous phrasings (significant at $p < .05$, Bonferroni corrected), in accordance with our predictions. A complementary finding is that undoing both *crash and poison* also got a high rating at the most general level of description, considering that was an option chosen by very few participants. Finally, *crash* was rated causally effective, independent of the level of description, as predicted by JDT.

Lastly, in order to examine the presence of the order effect predicted by Spellman (2005) and Roese (1997), the mean within-target Pearson correlation was calculated. A strong correlation between the ratings should be expected for the groups who listed the counterfactual factors first. However this values is of just .29 [df=70, p=.3]. When the causal judgements were presented first, correlation is .24 [df = 70, $p = .39$]. This constitutes very weak evidence to support Spellman et al.'s (2005) claims.

Probability ratings

Conditional probabilities differ significantly as a function of the target [$F(3, 67) = 155.83$, MSE = 1165.36, $p<.01$]. These judgements were kept consistent with Mandel's study, which means they were all set at the medium level. However, no specificity level effect was observed [$F(71) = .85$, $p = .43$]. The increase in the ΔP is the same as predicted by JDT, and against SPU [paired $t(71) = 5.20$, SED = 5.70, $p < .01$]. Surprisingly, the base rate of death given a *crime life* appears to be much higher in this study in comparison to Mandel's. The information is summarized in Table 3.

Table 3. Mean probability of protagonist death and ΔP.

	Mean	ΔP
Crime life	30.43	
Poison	66.03	26.4
Crash but not poison	70.71	
Crash and poison	83.60	12.17

Discussion

The dissociation between causal and counterfactual judgements predicted by JDT was not replicated in any of the levels of description. Participants and importance ratings were similarly distributed across types of judgements and varied across the specificity level of description. However, the modal response *crash* was chosen by most participants and consistently received high ratings, in agreement with the actuality principle. The main contrast between Mandel's and this experiment was observed in the ratings corresponding to the counterfactual task, specifically regarding *crime*. This target was rated as counterfactually efficient at the low level but dropped at the high level. Original JDT results could then be explained by the ambiguous objective of the counterfactual task. If the dissociation occurs due to a change in the task demand, it might be at least inappropriate to claim that such dissociation is due to a divergent function.

In summary, causal selection and counterfactual answers varied according to the level of description, more clearly in the contrast between high and low levels. Counterfactual answers were particularly more sensitive to the manipulation. Finally, no evidence in favour of an order effect was found, thereby ruling out Spellman's

explanation. This experiment shows how a minor modification in the instructions for generating causal and counterfactual judgements can have a tremendous effect on the focus of causal and counterfactual answers. Focus is led by the task demand and not by a functional difference, which of course means that dissociation could eventually occur, depending on the task's demands. The same information could be available for both types of judgement, perhaps as a causal model, even when the focus differs.

Experiment 2

In the previous experiment the dissociation between causal and counterfactual reasoning was not replicated. In fact, in most of the cases there was a match in focus between causal and counterfactual ratings, depending on the specificity of the description. However, there are at least two factors that can account for this finding that do not have a parallel in the original study. The first is the coding system, since a conjunctive category was included, and it accounted for an important proportion of participants. The second is the equivalence between the wording of the event description in the causal and the counterfactual task.

The second experiment examines the impact of the match of the wording between the causal and counterfactual tasks only, keeping the original coding. It is predicted that the findings of Experiment 1 will be replicated. More specifically, replication is predicted in the results of the medium level trend, where there was some ambiguity in the general objective of the counterfactual task. Order effects and the ambiguity of the objective of the counterfactual task are examined again.

Method

44 undergraduate psychology students at the University of Warwick were given course credit to participate in the experiment. The materials are the same as described for Experiment 1. The only difference is that all participants worked through causal and counterfactuals tasks that focus on the 'death of the main character', that is, the medium level in Experiment 1. As in Experiment 1, the order of the tasks was randomly counterbalanced. The coding was done according to the categories to Mandel's (2003b) original experiment, in order to make the results easily comparable.

Results

The proportion of participants per category and the pattern of counterfactual and causal ratings are summarized in Table 4. Similar numbers of participants chose *poison* and *crash* as causal and counterfactual targets. In the case of crime, more people consider it to be causally effective than effective in undoing the protagonist's death. However, the absolute number of participants who chose *crime* is significantly lower than the number of participants who chose either *poison* or *crash*.

Table 4. Percentage of participants and average ratings as a function of the task and specificity level.

Target	Counterfactual		Causal	
	%	M	%	M
Crime life	22	8.3	46	8.3
Poison	73	5.7	73	5.5
Crash	85	3.4	78	7.7

The ratings show a curious pattern that does not match the proportion of participants. Results were submitted to a repeated measures ANOVA. Target ratings changed depending on the task [$F(2, 81) = 5.33$, MSE = 60.52, $p < .05$], with *crash* rated as more important in the counterfactual than in the causal task [$t(43) = -4.5$, $p < .01$] although *crash* was chosen by fewer people in the counterfactual than in the causal task. The rating for *crash* is the lowest and significantly different from *poison* [$t(43) = -2.1$, $p < .05$] and *crime*[$t(43) = -4.6$, $p < .05$] for the counterfactual task. That is, crime was considered the most effective way of undoing the death of the protagonist, but was chosen by the fewest people. In the causal task, the ratings for all the targets were very similar, and planned comparisons revealed that crime and poison are considered similar in their causal effectiveness [$t(43) = -.2$, $p = .8$].

In summary, there appears to be a dissociation, not between judgements but between selection of a factor and its perceived effectiveness. *Crime life* was consistently rated as the best option (causal), but chosen by few people, whereas *crash* was chosen by most of the people in the counterfactual task, but perceived as ineffective in undoing the outcome.

Lastly, there was no order effect. Mean target correlation was calculated as a function of judgement order. Correlation was .19 [df =41,p =.23] when the causal task was presented first, and .01[df=41,p=.9] when the counterfactual task was presented first.

Probability ratings

Probability ratings significantly differ as a function of the target [$F(3, 43) = 58.64$, MSE=447, $p<.01$]. Again, the increase in the ΔP is as predicted by Mandel, and against Spellman, with the increase between *crime life* and *poison* being higher than the increase from crash [$t(43)=3.5$, SED=6.6 $p<.01$]. As before, base rate of death given *crime* is much higher in this study than in Mandel's. In fact, the size of the probability change in Exp. 1 and 2 is roughly half the size of the change in the original experiment.

Discussion

This experiment approximately replicated the results of the medium level of Experiment 1. No evidence was found in favour of JDT. The most important difference between these results compared to the medium level results in Experiment 1 is the proportion of people who chose *crime* as a cause, and the ratings attributed to *crash* across the causal and counterfactual tasks.

Probability ratings show the same trend as in Exp. 1 and in Mandel's experiments. However, an important difference

is the base rate probability of death given *crime life*. This is coherent with the high proportion of participants and high ratings attributed to *crime* in the causal task.

General Discussion

The experiments presented in this paper were designed with the objective of testing JDT and to determine whether causal selection and counterfactuals are actually related. In Experiment 1 the focus of causal and counterfactual judgements varied as a function of the description specificity of the target event. In Experiment 2 the impact of the instructions was isolated and confirmed. Both experiments failed to replicate the dissociation between causal and counterfactual reasoning and consequently cast doubt on the existence of a *functional* dissociation. Whereas the predicted match between causal and counterfactual reasoning did not emerge as clearly as expected, some insights may be drawn regarding both the systematicity of the effect and the shortcomings of JDT.

Participants consistently rated *crime life* as the most important factor in both causal and counterfactual tasks. The importance of this factor was lessened only when the tasks were described in the most specific manner (i.e. Exp. 1). In other words, the causal dependence on *crime life* was not considered only when the task referred to the specific way in which the main character died. This is also reflected in the base rate of death provided by the probability ratings. The objective and function of counterfactual reasoning are clearly task driven, but not necessarily in the form proposed by JDT (as generating ad hoc categories), but in actualizing causal contingencies of the situation as demanded by the task.

This calls into question the nature of what constitutes a sufficient cause and what is a 'preventor'. JDT proposes that counterfactual reasoning will focus on prevention, and preventors seem to be elements that enable a causal relation. Even accepting this hypothesis, it is the status of the preventor itself that is at stake when the dissociation occurs. The experiments show that what is considered a preventor (and a sufficient cause) depends on the particular description of the event and the instructions, which together constitute the task demand. That is, these notions are not possible to define independently of the context (Hart & Honoré, 1985), except in very limited settings. In the words of Lewis (1986), the events were *fragile*, since modifying their description radically altered the inferences made. Coincidence between causal and counterfactual reasoning thus crucially depends on using the same language to make the causal contingencies underlying the situation available to both types of reasoning. When a stable causal representation of the situation is made available to the reasoner, both causal and counterfactual queries are likely to converge (Sloman & Lagnado, 2005; Pearl, 2000).

References

Byrne, R. (2002). Mental models and counterfactual thoughts about what might have been. *Trends in Cognitive Sciences, 6*(10), 426-431.

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

Cheng, P., & Glymour, C. (1998). Causal mechanism and probability: A normative approach. In M. Oaksford & N. Chater (Ed.), *Rational models of cognition*. Oxford: OUP.

Collins, J. D., Hall, E. J., & Paul, L. (Ed). (2004). *Causation and counterfactuals*. Cambridge, Mass.: MIT Press.

Glymour, C., & Cooper, G. (1999). *Computation, causation, and discovery*. Menlo Park, Calif. Cambridge.; MIT Press.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the law*. Oxford: Clarendon Press.

Hilton, D., & Erb, H. (1996). Mental Models and Causal Explanation: Judgements of Probable Cause and Explanatory Relevance. *Thinking & Reasoning,2*, 273-308.

Hilton, D., Jaspars, J., & Clarke, D.(1990). Pragmatic conditional reasoning: Context and content effects on the interpretation of causal assertions. *Journal of Pragmatics, 14*(5), 791-812.

Kahneman, D., & Miller, D. T. (1986). Norm Theory: Comparing Reality to Its Alternatives. *Psychological Review, 93*(2), 136-153.

Kahneman, D., Slovic, P., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman & A. Tversky (Eds.), *Judgment under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press.

Lewis, D. (1986). *Counterfactuals*. Oxford: Basil Blackwell

Mandel, D. (2003a). Effect of counterfactual and factual thinking on causal judgements. *Thinking & Reasoning, 9*(3), 245-265.

Mandel, D. (2003b). Judgment dissociation theory: an analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of experimental psychology. General, 132*(3), 419-434.

Mandel, D., & Lehman, D. R. (1996). Counterfactual Thinking and Ascriptions of Cause and Preventability. *Journal of personality and social psychology., 71*(3), 450.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press.

Roese, N. (1997). Counterfactual thinking. *Psychological bulletin, 121*(1), 133-148.

Sloman, S. A., & Lagnado, D. A. (2005). Do We "do"? *Cognitive Science, 29*(1), 5-39.

Spellman, B. A. (1997). Crediting Causality. *Journal of experimental psychology. General, 126*(4), 323.

Spellman, B. A., Kincannon, A., & Stose, S. (2005). The relation between counterfactual and causal reasoning. In D. Mandel, D. Hilton & P. Catellani (Ed), *The psychology of counterfactual thinking*. London: Routledge.

Verschueren, N., Schaeken, W., & Schroyens, W. (2006). Necessity and sufficiency in abstract conditional reasoning. *European Journal of Cognitive Psychology, 18*(2), 255-276.

Wells, G., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56*(22),161 - 169.