Causal Selection and Counterfactual Reasoning

Selección Causal y Razonamiento Contrafactual

Seleção Causal e Raciocínio Contrafactual

WILLIAM JIMÉNEZ-LEAL
Universidad de los Andes, Bogotá, Colombia

Abstract

In this paper I defend the view that counterfactual thinking depends on our causal representation of the world, and in this sense, I argue that causal and counterfactual reasoning are tightly linked. I offer some criticisms and experimental evidence against Mandel's judgement dissociation theory (Mandel, 2003b), which claims functional independence between the process of causal selection and counterfactual reasoning in the context of causal selection. In the experiments described, I manipulated some elements of the semantics of the task to show the cases in which dissociation between causal and counterfactual reasoning does not occur. In Experiment 1, the level of description of the target event is manipulated in a list generation and rating task. Experiment 2 replicates Experiment 1 findings using an alternative coding system, whereas Experiment 3 does the same using an alternative answer format. The results of the experiments support the picture of causal understanding proposed by the causal mental models.

Keywords: counterfactual reasoning, causality, judgement dissociation theory

Resumen

El trabajo defiende la posición según la cual el pensamiento contrafactual depende de nuestra representación causal del mundo y, en este sentido, argumenta que existe una estrecha relación entre el razonamiento causal y el contrafactual. Se lleva a cabo una crítica a la teoría de la disociación de juicios de Mandel (Mandel, 2003b), que defiende la independencia funcional entre el proceso de selección causal y el razonamiento contrafactual en el contexto de la selección causal. En los experimentos realizados se manipularon algunos elementos de la semántica de la tarea con el fin de ilustrar aquellos casos en los que no se da la disociación entre el razonamiento causal y el contrafactual. En el Experimento 1, el nivel de descripción del evento objetivo se manipuló en una tarea de generación de listas y evaluación. El Experimento 2 replicó los hallazgos del Experimento 1 utilizando un sistema de codificación alternativo, mientras que el Experimento 3 realizó lo mismo utilizando un formato de respuesta alternativo. Los resultados de los experimentos apoyan la concepción del entendimiento causal propuesta por los modelos mentales causales.

Palabras clave: razonamiento contrafactual, causalidad, teoría de la disociación de juicios.

Resumo

O trabalho defende a posição segundo a qual o pensamento contrafactual depende de nossa representação causal do mundo e, nesse sentido, argumenta que existe uma estreita relação entre o raciocínio causal e o contrafactual. Realiza-se uma crítica da teoria da dissociação de juízos de Mandel (Mandel, 2003b), que defende a independência funcional entre o processo de seleção causal e o raciocínio contrafactual no contexto da seleção causal. Nos experimentos realizados, manipularam-se alguns elementos da semântica da tarefa com o objetivo de ilustrar aqueles casos com os quais não se dá a dissociação entre o racionamento causal e o contrafactual. No Experimento 1, o nível de descrição do evento objetivo se manipulou em uma tarefa de geração de listas e avaliação. O Experimento 2 repetiu as descobertas do Experimento 1 ao utilizar um sistema de codificação alternativo, enquanto o Experimento 3 realizou o mesmo ao utilizar um formato de resposta alternativo. Os resultados dos experimentos apoiam a concepção do entendimento causal proposta pelos modelos mentais causais.

Palavras-chave: raciocínio contrafactual, causalidade, teoria da dissociação de juízos.

Correspondence concerning this article should be addressed to William Jiménez-Leal, e-mail: w.jimenezleal@uniandes.edu.co. Department of Psychology, Universidad de los Andes, Cr. 1 No. 18A-10, Building Franco, 2nd floor, Bogotá, Colombia.

ARTÍCULO DE INVESTIGACIÓN CIENTÍFICA
RECIBIDO: 1 MARZO DE 2013 - ACEPTADO: 4 DE ABRIL DE 2013

* Special thanks to Nick Chater and an anonymous reviewer for their relevant comments.

IN APRIL 2003, a team of BBC journalists was accompanying a Kurdish and American convoy southeast of Mosul, as part of their coverage of the Iraq war. The convoy was bombed by mistake by an American jet, killing 15 soldiers and injuring several people, including some journalists. The producer´s mother happened to call to wish him a happy birthday just a few seconds before the bomb hit, and as he stepped away holding his phone up ("this is the sound of freedom" he said to her), the bomb was dropped. He would later say that if she had not rung him, he would have been killed ("Friendly fire' hits Kurdish convoy", 2003).

Examples abound to illustrate the pervasiveness of counterfactual thinking and its importance in our mental life. Still, its cognitive function remains controversial, particularly when related to causal reasoning. In the example presented, the reader is compelled to accept the truth of the statement, but did his mother's call actually save the producer's life? That is to say, is the call the cause of the event? And if so, what can we say about the cases where counterfactual reasoning helps us select the cause of an event?

Causal selection represents a psychological puzzle: How, amongst the myriad of factors that are present in a given situation, do people select those that are considered causal? Research in psychology has traditionally focused on situations with very simple causal structures (Kelley, 1983) or cases that can be described with correlation information, where the candidate cause is already given (Cheng, 1997). Einhorn and Hogarth (1986) propose that in evaluating causality people attend to several cues that can offer information on the events' relationship: contiguity, covariation, and temporal order are some of them. More recently, researchers have identified that people also prefer to attribute causality when mechanism information is available (Ahn, Kalish, Medin, & Gelman, 1995) and when a human action can be identified in the causal sequence (McClure, Hilton, & Sutton, 2007). However, the role of counterfactual reasoning in causal selection is still widely debated.

A possibility considered in philosophy (Lewis, 1973) is that causality can actually be reduced to a counterfactual dependence, which in turn can lead one to consider that causal understanding depends on counterfactual understanding. From a psychological perspective, this would be equivalent to proposing that understanding that X causes Y depends on being able to conceive an alternative in which X does not cause Y, in other words, to entertain the counterfactual thought that "if X had not happened, Y would not have occurred". This idea has been dubbed the *counterfactual process view of causal thinking* (Hoerl, McCormack, & Beck, 2011). An alternative view is held by Mandel (2003b, 2011), who proposed judgement dissociation theory (JDT) to explain why in some situations causal and counterfactual reasoning focus on different factors. For example, one may consider that a spark is the cause of a particular fire, but also think that if Mary had not left the candle burning, the fire would not have occurred.

Both proposals represent opposite ends of a spectrum of judgement on causal selection and are impossible to reconcile. In what follows, I claim that counterfactual reasoning does have a key role in causal selection (*contra* JDT) but based on the idea of Bayesian causal models (Sloman, 2005). Causal and counterfactual reasoning can potentially focus on different factors, but this fact does not imply that they are functionally distinguishable. At best, what this fact implies is that both types of reasoning are context sensitive. I assert that the causal representation of the world is a conceptual primitive, and that counterfactual reasoning can help uncover this representation. I will first outline the main tenets of JDT and show some problems with this theory. I propose an alternative view on the link between causal and counterfactual reasoning based on how this link is conceived in the context of Bayesian causal models. The experiments

reported here are based on the idea that counterfactual reasoning can help the process of causal selection when the relevant events are identified in the same level of description, a principle implied by Bayes nets.

Judgement Dissociation Theory and Counterfactual Reasoning

Mandel (2003b) puts forward JDT to conceptualize the apparent independence between the process of causal selection and counterfactual reasoning. The operation of JDT is based on two principles, actuality and substitution. According to the actuality principle, causal selection is based on identifying sufficient antecedents for the actual outcome 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, that operate independently, are based on different information (Mandel & Lehman, 1996), have different targets and do not influence each other (Mandel, 2003a).

The alternative, currently out of favour, is the counterfactual process view of causal thinking or counterfactual simulation theory. This is a loose of group of proposals (N'gbala & Branscombe, 1995; Petrocelli, Percy, Sherman, & Tormala, 2011; Wells & Gavanski, 1989; Wells, Taylor, & Turtle, 1987) that can be roughly characterised by the idea that people identify the causes of an event by performing a mental simulation of the negation of the candidate cause.

Mandel (2003a, 2003b) points out some undesirable consequences of this perspective: the idea that causation is a relation of necessity, not sufficiency; and that understanding counterfactual statements is equivalent to understanding causal statements. Clearly both are wrong and thus justify rejecting the simulation approaches in favour of JDT.

Mandel uses cases of pre-emption to analyse and criticise counterfactual simulation accounts, echoing the discussion in philosophy (Collins, Hall, & Paul, 2004). Take as an example the case of Suzy and Billy. They both throw a rock at a bottle at the same time, but Suzy's rock hits the bottle first, breaking it. In this case, a counterfactual simulation does not allow determining the cause since both counterfactual simulations of the absence of the candidate causes fail to undo the effect. Consider statements (1) and (2), as representing the corresponding simulation.

1. If Suzy had not thrown the rock, the bottle would not have broken.
2. If Billy had not thrown the rock, the bottle would not have broken.

In both cases the counterfactuals are false, indicating that eliminating the candidate cause does not eliminate the effect, therefore the simulation fails to identify the cause.

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 counterfactual simulation. The substitution principle predicts that people will focus on preventors during counterfactual resasoning, elements that are also out of the scope of the rival theories.

Problems with JDT

There are some problems with the way JDT is specified. In what follows, I point out three: the notion of sufficiency and preventor, its interpretation of counterfactual simulation, and the contraposition of mechanism and dependence information. The key concepts of "sufficiency" and "preventor" are underspecified. There is no broad agreement on how to define them, and even 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, Schaeken, & Schroyens, 2006). Some researchers have convincingly argued in favour of a contextual definition of sufficiency and necessity (See Hart & Honoré, 1959; Hilton & Erb, 1996; Hilton, Jaspars, & Clarke, 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 characterisation of sufficiency and necessity under the circumstances are left unexplained.

On the other hand, Mandel's concept of counterfactual reasoning is based on an illicit generalisation of the idea of the simulation heuristic (Kahneman, Slovic, & Tversky, 1982). Within Norm Theory (Kahneman & Miller, 1986), the simulation heuristic operates as a post facto reaction that is motivationally relevant. Therefore, not all counterfactual simulations aim to identify causes (see Petrocelli et al., 2011). Furthermore, even when counterfactual reasoning is required to answer a causal question, the interpretation of the question plays a key role in determining the norm against which to compare. By taking the operation of the simulation heuristic out of context, as proposed by Norm Theory, Mandel equates its operation with all counterfactual reasoning. He fails to acknowledge that counterfactual contingencies are actualised by the demands of the task, and made available for causal selection.

Finally, Mandel wrongly opposes causal mechanism information to probabilistic information in their importance for causality attribution. Causal selection can rely on either of these sources, and what is more, in many cases they provide equivalent information (Cheng & Glymour, 1998). Again, this could be due to the lack of the conversational dynamics. Where Mandel wants to describe different questions (how vs. what) he describes different sources of information. The difference lies in the task demand, not in the information itself.

There are, however, good reasons why the object of counterfactual reasoning can be mismatched with the object of causal reasoning. I believe the key notions that JDT lacks are those of sensitive causation (Woodward, 2006) and modal fragility of an event (Lewis, 1987). A causal relationship is said to be sensitive to external factors when it holds in the actual circumstances but would not continue to hold in circumstances that depart, even slightly, from the actual. An insensitive causal claim holds in the actual circumstances and would continue to hold across a range of changes from the actual circumstances. To use one of Lewis' examples, shooting a victim at point blank through the heart is a case of causal insensitive causation since small variations will not alter the outcome, whereas running into an old friend because you left your house late in the morning is a case of sensitive causation because it could have easily happened in a different way (for a similar position see Menzies, 2011). Similarly, the causal *relata* might be modally fragile, meaning that an event occurring at a different time and in a different manner would be considered a modally different event and thus would be subject to a different set of counterfactual dependencies. Thus, death by a heart attack implies a different counterfactual history than death by gunshot, and both are modally different than simply describing an event as an unqualified death.

I maintain that counterfactual reasoning only allows singling out causal factors in cases of either insensitive causation or where the *relata* are modally robust. Psychologically, this simply means that a counterfactual does not identify the cause if the causal events can be instantiated in several alternative ways, and if the causal link itself can be instantiated in several ways. In the end, this comes down to the way a causal relationship is represented in the context of explanation, which in turn depends on the demands of the task or situation. JDT can

explain cases of dissociation, clearly, but not the matches, because it lacks the theoretical tools to handle changes in the way the representation of the causal *relata* are specified.

There is also evidence that points to integration, rather than to dissociation of causal and counterfactual reasoning. Byrne (2002) has pointed out that a good guide for understanding counterfactual reasoning is to explore the representation of the factual possibilities associated with it. The appropriate modelling of the causal structure might help to understand how causal and counterfactual reasoning relate, and at the same time to provide a normative framework for studying counterfactuals (Sloman & Lagnado, 2005). This framework can easily incorporate the notions of sensitive/insensitive causation and fragile/robust events.

An Alternative

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 of an event, most people do not change their beliefs about the state of the normal causes of the event" (p. 27). That is, the causal structure against which counterfactuals are judged is kept stable. Moreover, Sloman and Lagnado (2005) 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, that is, causal inference is determined by counterfactually altering the values of a variable, as part of a causal network.

How is it possible then to reconcile these findings with Mandel's results? Spellman and colleagues (Spellman, Kincannon, & Stose, 2005) proposed that dissociation is an order effect. They assert that in cases in which the same participants have to complete causal and counterfactual tasks the following configurations are possible: When subjects are asked first to generate counterfactuals, that information regarding the counterfactuals then becomes available for causal evaluation, making it more probable to affect the performance in the causal task. However, when the causal task is performed first, this does not affect mutations performed afterwards. However, contrary to Spellman et al. (2005), Mandel (2003a, 2003b) did not find any order effects. This explanation is considered in the following experiments.

An alternative explanation is that the mismatch observed by Mandel occurs because causal and counterfactual queries are not usually specified in the same way. In fact, Mandel's counterfactual probes always refer to undoing the "outcome of a situation" whereas causal questions refer to a particular event (Mandel, 2003a, p. 423). In other words, causal queries relate to insensitive causal relationships, whereas counterfactuals are centred on sensitive causal ones. Similarly, the instructions for the probability ratings he requested were not matched. 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 structure is clear (causally insensitive), it is feasible that counterfactual and causal tasks can have the same targets (they are modally robust). Causal queries convey cues that somehow specify the content of the causal answer(s) they are intended to receive. Thus, a question about what is 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 to contrast the results with Mandel's.

Experiment 1

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

Method

Participants

Seventy-two undergraduate students (43 female and 29 male) from different programs at the University of Warwick took part in the experiment in exchange for payment (£ 3.50). The mean age was 20.2 (SD=2.4) and all participants successfully completed the task.

Materials and Procedure

Participants were tested individually in a cubicle with a computer-based experiment. The selected stimuli were presented to participants on a computer screen using a program written by the author in the Delphi programming language (Texeira & Pacheco, 2001). Participants worked on the task at their own pace and all of them completed the tasks requested.

The complete display included eight screenshots and followed the structure of Mandel's experiments. The first screenshot contained the general instructions, where it was emphasized that they would have the opportunity to read a vignette only once and then they would be asked questions about it. The vignette was presented in the second screenshot, where a criminal falls prey to two assassination attempts. Briefly, 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.

Once the participants had read the scenario, they proceeded to complete the causal, the counterfactual and the probability 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.

For the causal task, participants were asked to list up to four factors that they "regard as causes of the 'event'". In the next screenshot they were asked to rate the importance of these factors (from 0 to 10): "Now please rate the importance of each factor you listed with regard to causality on a scale of 0-10 where 0 'not at all causal' and 10 'totally the cause'".

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 from 0 to 10 "how likely those alterations would have been in changing the 'event': 'Please rate the importance of each of the changes you listed with regard to how likely that change would have been in altering the event on a scale of 0-10 where 0 'not at all a good way to undo the event' death and 10 'absolutely the best way to undo the event'".

The description of the "event" varied in three levels and for each level both causal and counterfactual tasks were matched. Three levels of specificity were defined for the event description: For the first or low level, the questions were about the "outcome of the situation". For the second or medium level, the judgements required were about the "death of the main character". For the third or high level, participants were asked about the "death of the main character due to the fatal burns".

Participants were randomly assigned to one of the three versions defined by the specificity level of event description. Notice that the main difference between these tasks and Mandel's is the variation of the event description and the matching of the description across tasks. In the counterfactual task, Mandel (2003b) asked people about ways the "story could be changed so that the outcome would have been different" and

then rate how effective these were in undoing the "character's *premature* death". In the causal task the questions were simply about the death of the character.

In the probability task, participants were asked to estimate four probabilities (from 1 to 100) for the outcome, given four conditions, defined by the presence/absence of the actions of the assassins:

1. None of them occurring.
2. Poisoning but not car run off the road occurring.
3. Car runoff road but not poisoning occurring.
4. Both of them occurring.

For example, the sentence corresponding to condition 1 was:

"What is the probability of the victim dying given that neither Mr. Vincent added poison to Mr. Wallace's drink nor Mr. Bruce pushed Mr. Wallace's car into a ravine."

Coding of Causal and Counterfactual Listings

The coding was done according to the categories of interest: crime life, poison, crash and other. Two additional elements were recorded: first, in the case of the counterfactuals, whether or not the manipulation actually undid the death of the main character; second, the frequency with which participants mentioned the conjunction of any of the targets (e.g., poison and crash). Each participant provided at least one answer for the causal and the counterfactual listings. Coding was performed by the author and by an independent coder. Inter coder agreement was 88% (Raw agreement index. 498/549=90% Overall. For causal answers 92%, for counterfactual 88%. Kappa coefficients are .92, $p<.01$ and .96, $p<.01$ respectively.).

Results

Importance ratings were computed by dummy coding participants' causal and counterfactual listings as 0 if the target was absent and as 1 if the target was present. Then each of the answers was weighted on the basis of the importance rating given to it and averaged if any target was selected more than once. Mean counterfactual and causality scores for each participant were then calculated by averaging the sum of the scores by the number of answers.

In Mandel's study, listings and ratings followed the same pattern, so the analysis focused on the ratings. This finding was not replicated. In what follows, I first present the overview of response frequencies followed by the modal responses and the importance ratings analyses, and finish with the probability ratings.

Proportions of Answers

Participants produced 269 and 289 answers for the causal and counterfactual listings. The difference is not significant [$\chi^2(1, n=549)=0.22$, $p=0.63$]. Table 1 summarizes the overall selections for both tasks. It shows first the percentage of participants listing each target and then the mean importance ratings for each one of them. Contrary to JDT predictions, overall crime life

Table 1
Overall Frequencies and Percentages of Responses by Target as a Function of the Task

Target	Counterfactual		Causal	
	n	%	*n*	%
Crime life	28	9%	62	23%
Poison	75	26%	51	19%
Crash	107	37%	81	30%
Crash and Poison	58	20%	8	3%
Other	29	10%	67	25%
Total	289	100%	269	100%

was not the preferred modal answer for counterfactual task, nor crash for the causal task.

There is a significant difference between the frequencies of people who chose a counterfactual target [$X^2(4, n=289)=86.06, p<.01$]. Overall, most of the answers focused on undoing the crash to undo the event, followed by poison and the combination of both. There is also a significant difference among causal targets [$X^2(4, n=269)=57.37, p<.01$]. There is a large number of participants who manipulated both the car accident and the poisoning in the counterfactual but not in the causal task. The highest number of responses was given in the category crash, for both types of judgements. Surprisingly, crime life was chosen more frequently as the target in causal compared to counterfactual listings. There was also a large difference between the number of responses that do not fall into any category in the causal and the counterfactual task: Whereas this accounts for just 10% of the counterfactual answers, it is relevant to 25% of the causal answers.

A log linear analysis was conducted to test for differences across categories according to the level. It included specificity level (3), judgement type (2), and target (3)[1]. The three way log linear analysis produced a final model that retained the specificity level, target and the judgement type and target interactions (but not the specificity level x judgement type interaction). The likelihood of this model was [$\chi2(6)=3.75, p=.71$]. The interaction between the specificity level and the target was significant [$\chi2(4)=15.16, p<.01$], which indicates that the number of responses for target differ across the specificity level. In particular, the highest difference between crash and crime life occurs in the high specificity level, independent of the judgement type, indicating, first, match in the importance attributed to this factor in both causal and counterfactual judgements, and, second, that its importance increases with the specificity of the question.

The interaction between the type of judgement and the target was also significant [$\chi2 (2)=24.91, p<.01$]. It can be seen that the percentage of answers for crime life was higher for causal than for counterfactual judgements. The odds ratio suggests that a participant is 3.95 more likely to judge crime life causally relevant than counterfactually efficient in undoing the outcome of the story compared to the other targets (collapsing the other 2 categories). This clearly diverges from Mandel's results, where the frequency of judgements on crime life was clearly higher for counterfactual than for causal judgements. Analysis of the proportion of participants per target, not answers, was also performed. The same results were observed (Model: $\chi2 (6)=4.43. p=.61$ and the same effects (specificity level x target and judgment x target).

Proportions of Participants and Importance Ratings

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

The crash and poison target follows a similar pattern for both the causal and the counterfactual tasks, with the lowest proportion of people choosing them in the low level, and the highest in the high level. That is, a similar number of people considered these targets to be

1 It is important to bear in mind that the log linear analysis was performed on the distribution of the total number of answers across categories, and not the number of respondents in each category. Log linear analysis was also run with 4 targets, showing no significant difference.

the cause with independence of the level [$X^2(2, n=5)=0.4$, $p=.8$].The highest proportion of people in the counterfactual task corresponds to the high specificity level, where it was chosen by 33% of the participants in that level. However, the difference between the levels is not significant [$X^2(2, n=17)=1.41$, $p=.48$].

Overall, it can be seen that proportions of people choosing a target are fairly similar across the tasks, and that the higher number of people for both tasks is concentrated around the poison and crash targets. The distribution of the percentage of participants roughly mirrors the overall distributions of answers. There are no significant differences between the number of participants who chose poison or crash as cause of the events in a test across levels (poison [$X^2(2, n=45)=2.4$, $p=.29$]; crash [$X^2(2, n=52)=.05$, $p=.97$]). In the counterfactual task the same pattern emerges (poison [$X^2(2, n=51)=1.1$, $p=.59$]; crash [$X^2(2, n=53)=0.6$, $p=.71$]).

Table 2
Percentage of Participants and Average Ratings as a Function of the Task and Specificity Level (Experiment 1)

		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

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)). The results show significant effect of judgement type [$F(1, 69)=15.99$, MSE=93.67, $p<.01$], with a larger marginal mean for causal judgements, and target [$F(3, 67)=17.30$, MSE=265.83, $p<.01$]. There is also interaction between target and judgement type, [$F(3, 67)=8.60$, MSE=78.20, $p<.01$] parallel to the results found with the proportions of answers. 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 marginal means of the specificity level did not differ significantly, but the ratings assigned to the targets did vary as a function of the level.

Planned comparisons revealed that for the causal judgements there was a significant difference between crash and poison [$t(71)=-3.19$, $p<.01$, crash higher than poison], but not between crime life and the others [poison $t(71)=.56$, $p=.57$, crash $t(71)=1.64$, $p=0.1$]. For the counterfactual judgements the opposite pattern arises: Crash and poison are not significantly different [$t(67)=.36$, $p=0.71$], but crime life is different from poison [$t(67)=-3.42$, $p<.01$] and crash [$t(67)=-3.45$, $p<.01$] (both crash and poison are significantly higher). Remember that there was no interaction between specificity level judgement types, but there was with target.

The ANOVA was also run dropping the fourth target (crash and poison) as a way of keeping the coding as comparable as possible to Mandel's analysis. The results are almost identical (main effect of target and judgement type, plus interaction between them; interaction between specificity level and target, but not target).

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). A complementary finding is that undoing both the crash and taking the poison also got a high rating at the most general level of description, considering that was an option chosen by very few participants. For the causal task, in the high level, more people consider the combination of both important factors (crash and poison) to be the cause. This category had a similarly high rating compared with crime life in the same level. Finally, crash was rated causally effective independently of the level of description.

Lastly, in order to examine the presence of the order effect, the mean within-target Pearson correlation was calculated. When the counterfactual judgements were presented first, the results show correlation values of .29 [df=70, p=.3]. When the causal judgements were presented first, correlation results were .24 [df=70, p=.09] (drop to 0.25 and 0.20 if the fourth target is not included). The results are then evidence for the absence of order effects.

Effectiveness of Counterfactual Responses

Effectiveness of the counterfactual responses can be considered by examining the proportion of answers that actually undid the intended event. In this case, the proportion of counterfactual responses that failed to undo the death of the protagonist reached 30%. They represent 30%, 6%, and 1% of the low, medium and high description specificity factors, respectively, and the participants were not equally distributed across the levels of specificity [$\chi^2(2, n=27)=5.67$, $p<.05$], with more answers failing to undo the death of the protagonist in the low specificity level. This indicates that the objective of the counterfactual task was not interpreted as equivalent across the specificity levels. The information is summarized in Table 3.

When the level of specificity was high, the counterfactuals generated did not necessarily undo the death of the character. For example, some of the participants' answers in the low specificity description explicitly mentioned that the protagonist still dies (e.g., "He stops the car and gets out but the van hits him" and "Someone else kills him before").

Table 3
Number of Counterfactual Modifications that Failed to Undo the Death of the Protagonist by Target (Experiment 1)

	Crime life	Poison	Crash	Other	Both
Low	0	15	14	7	2
Medium	0	8	10	6	1
High	0	7	7	4	0
Total	0	30	34	17	3

It can be seen that none of the judgements that focused on crime life failed to undo his death, in clear contrast with poison and crash. Most of these modifications involved interaction between them. In another example: "he could have survived the car accident and still die poisoned in hospital minutes later". Effectively, these are changes in the outcome, as defined in Mandel's original experiment, which failed to change the death of the protagonist. In other words, these are counterfactuals changes to an insensitive causal link.

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 consistent with Mandel´s prediction. What is striking is that the base rate of death given a life of crime appears to be much higher in this study in comparison to Mandel's. The information for this experiment is summarized in Table 4.

Table 4
Mean Estimated Probability of Protagonist Death and Probability Change (Experiment 1)

	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. The proportion of answers, participants and importance ratings was similarly distributed across types of judgements and varied across the specificity level of description. It was also found that the counterfactual the task had an ambiguous objective, which may possibly explain the original JDT results.

The modal responses varied as a function of target, although not as precisely as it has been expected. The main contrast between this experiment and Mandel's was observed in the proportions and ratings corresponding to the counterfactual task, specifically regarding the target crime. This target was rated as counterfactually efficient in the low level and significantly dropped in the high level. This stands in clear contrast with JDT predictions. In summary, cause selection and counterfactual answers varied according to the level of description, more clearly in the contrast between high and low-medium levels. Counterfactual answers were particularly more sensitive to the manipulation.

More importantly, modal selection of target did not necessarily imply undoing of the target event. Counterfactuals generated in the low specificity level did not necessarily undo the death of the main character of the story. Mandel asked people about ways in which the "story could be changed so that the outcome would have been different". It is possible to modify the story even if the main character had died anyway. The original classification by target (criminal life, explosion, and poison) takes for granted that people introduce changes to undo the death of the protagonist. Finally, no evidence in favour of an order effect was found, therefore ruling out the possibility of the dissociation due to an order effect.

This experiment showed how a simple modification in the instructions for generating causal and counterfactual judgements can have a tremendous effect on the overall objective of a counterfactual task, and 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 can occur depending on the context's demands.

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 clear 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 one is clearly the coding system, since a conjunctive category was included, and it accounted for an important proportion of participant responses. The second one is the equivalence between the wording of the event description in the causal and the counterfactual tasks.

The second experiment examines the impact of the match of the wording between the causal and counterfactual tasks, keeping the coding suggested by Mandel. It is predicted that the findings of Experiment 1 will be replicated, more specifically, the medium level findings, 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 were examined again.

Method

Participants

Forty four undergraduate psychology students (30 female and 14 male) were given course credit to participate in the experiment. Their mean age was 19.8 (SD=1.3) and all of them successfully completed the task.

Materials and Procedure

The materials are the same as described for Experiment 1. The only difference is that all participants work 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 of a priori interest, as in Mandel's (2003b) original experiment, in order to make the results more directly comparable. As a measure of the ambiguity of the task, it was recorded whether or not the counterfactual manipulation actually undid the death of the main character. Answers were again coded by the author and the same coder of the first experiment.

Results

The results will be presented in an order similar to Experiment 1. I first present a general overview of the frequencies of responses. Then, the proportion of participants and the mean ratings per target are summarized and compared with the previous experiment, as well as the order effects. This is followed by the ambiguity measure of the counterfactual task objective. Finally the probability ratings are examined.

Proportions of Answers

There is no significant difference between the number of overall causal and counterfactual responses [$\chi^2(1, n=278)=0.7, p=.4$]. However, the number of answers differ depending on the target [$X^2(3, n=278)=12.51, p<.05$], with fewer responses attributed to crime life in both tasks (see Table 5).

Table 5
Overall Frequencies and Percentages of Responses by Target as a Function of the Task (Experiment 2)

Target	Counterfactual		Causal	
	n	%	*n*	%
Crime life	10	7%	28	19%
Poison	50	34%	35	24%
Crash	56	38%	47	32%
Other	30	21%	22	15%
Total	146	100%	132	100%

There are no significant differences across tasks, with both types of judgements exhibiting roughly the same proportions of responses per target. The exception is crime life, with more responses in the causal than in the counterfactual category [$\chi^2(1, n=38)=8.4, p<.01$].

Proportions of Participants and Importance Ratings

The proportion of participants per category and the pattern of counterfactual and causal ratings are summarized in Table 6. A similar number of participants chose poison and crash as causal and counterfactual targets. In the case of crime, more people considered it to be causally effective than effective in undoing the protagonist death. However, the absolute number of participants who chose crime is significantly lower than the number of participants who chose either poison or crash. This pattern closely reflects the proportions of answers examined before in Experiment 1.

Table 6
Percentage of Participants and Average Ratings as a Function of the Task and Specificity Level (Experiment 2)

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 the 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 [$t(43)=-.2$, $p=.8$] but different from poison [$t(43)=-3.19$, $p<.01$]. In summary, there appears to be dissociation, but in a direction contrary to that predicted by JDT: People chose the "necessary" elements as causally effective and the sufficient elements as counterfactually effective. The ratings for the target differed as a function of the task. In particular, ratings for crash, complementing an opposite trend observed with the proportion of participants who chose each target. Additionally, poison is not considered as causally effective as crime and crash.

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.

Effectiveness of Counterfactual Responses

These responses were distributed among 11% of the participants. These participants produced at least one counterfactual modification that did not undo the death of the main character. Crash was the most common target associated with a failed counterfactual, which fits the picture of being the most frequently chosen counterfactual target, but with the lowest mean rating. This result is coherent with findings in Experiment 1, where the counterfactual responses at the medium level specificity condition were very specific descriptions of the target of the counterfactual.

Probability Ratings

Probabilities ratings significantly differ as a function of the target [$F(3, 43)=58.64$, MSE=447, $p<.01$].The ratings are summarized in Table 7. Again, the increase in the ΔP is as predicted by Mandel 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 in the first experiment, base rate of death given a life of crime appears to be much higher in this study in comparison to Mandel's. In fact, the size of the change in probability in Experiment 1 and 2 is roughly half the size of the change in the original experiment.

Table 7
Mean Estimated Probability of Protagonist Death and Probability Change (Experiment 2)

Target	Mean	ΔP
Crime life	38.43	
Poison	75.85	37.42
Crash but not poison	83.90	
Crash and poison	97.15	13.25

Discussion

This experiment approximately replicated the results of the medium level of Experiment 1. No evidence of target dissociation between causal and counterfactual reasoning was found. 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. Crash is coherently considered a cause by examining both proportions and ratings. However, it was also chosen by a large proportion of

participants who nonetheless recognized that modifying the crash is not enough to undo the death of the main character. It shows dissociation, not as a function of the judgements, but of the evaluation considered in a causal network and its relative weight in producing an effect. The attributions to crime were again higher for the causal than for the counterfactual task, in line with the probability ratings.

Probability ratings exhibit the same pattern as in Experiment 1 and in Mandel's experiment. However, a very important difference consists in the base rate probability of dying given that the protagonist is a gangster. In both Experiments 1 and 2, the base rate probabilities are nearly double the base rate in Mandel's experiments. This is coherent with the high proportion of participants and ratings attributed in the causal task to crime life.

In both Experiments 1 and 2 the probability ratings do follow the predictions of JDT. However, given the rest of the results, it makes sense to affirm that this is not due to a functional dissociation. One possibility is that the dissociation does not emerge due to problems with the coding. Many of the cases classified as other might have altered the answer and ratings distributions if switched to a different category. This hypothesis is tested in Experiment 3 by using closed answers.

Experiment 3

Results of the second experiment showed that the dissociation between causal and counterfactual reasoning is explained by the match between the descriptions of the events. The pattern of the modal responses exhibits a trend similar to the medium specificity level in Experiment 1. That is, the match between causal and counterfactual reasoning, at least in the modal responses, is not that clear.

It is possible that the methodology used to collect the data has some effect on this result. Previous research has either used closed answers or listing generation (Mandel & Lehman, 1996; N'gbala & Branscombe, 1995; Wells & Gavanski, 1989), but not both. Mandel (2003b) has rightly argued that the demand in each case is different, and that part of the success of JDT is being able to predict performance across differential demand. The differential demand consists of the availability of the options for listings, since the generation of answers depends more directly on what the focus of attention is. However, what is not taken into account is the possible interference of one task over the other. In all of Mandel's experiments, as well as in Experiments 1 and 2, participants were first asked to list factors and then to rate them. It is probable that the demand on focus to list the factors had altered the attributed importance. The use of predefined categories would make the relevant causal categories obvious, and consequently alleviate any contextual influences.

Hence, the use of predefined categories is implemented in this experiment. It is predicted that the match between causal and counterfactual judgements will be more clearly replicated if the task demands are independent. Otherwise, the pattern should change. In this experiment the causal and counterfactual tasks consist only of the ratings of the content categories of a priori interest. No listings or probability ratings were requested from the participants.

Method

Participants

Forty one volunteer undergraduate students (25 female and 16 male) from several programs at the University of Warwick were paid £2 to complete Experiment 3.

Materials and Procedure

The method was generally the same as described for Experiment 2. The causal and counterfactual tasks consisted exclusively in rating the factors provided so the wording was adjusted

slightly to reflect this, while keeping the match on focus between tasks unchanged. No coding was necessary since the closed answer method was adopted. Probability ratings were not requested, making the overall duration of the experiment shorter.

Results

The mean counterfactual ratings for crime life, poison and crash were 8.1, 5.9, and 4.9 respectively, while the mean causal ratings were 7.9, 6.1 and 6 respectively; thus, rating results of Experiment 2 were closely replicated, thereby not providing support to the idea of task interference.

Results were submitted to a repeated measures ANOVA. Ratings varied by target [$F(2, 80)=17.5$., MSE=119.35, $p<.01$], but target ratings did not vary depending on the task [$F(2, 80)=3.8$, MSE=29.5.52, $p=.06$]. Overall, crime life was considered more important than the other targets (planned contrasts: crime vs. poison [$F(1, 40)=24.33$., MSE=146.5, $p<.01$]; crime vs. crash [$F(1, 40)=36$, MSE=206.4, $p<.01$]. The pattern is the same within each type of judgement. In the counterfactual task, crime is significantly higher than poison [$t(40)=3.8$, $p<.05$] and [$t(40)=7.3$, $p<.05$] and crash, the same occurring in the causal task ([$t(40)=4.3$, $p<.05$] and [$t(40)=2.9$, $p<.05$]).

Discussion

The general pattern of results from Experiment 2 was replicated in Experiment 3. No evidence in favour of JDT was found, nor evidence in favour of a task interference hypothesis that could explain the results found regarding importance ratings across the experiments. Again, crime life was rated as a far more important factor than poison and crash, which received similar ratings. However, crash ratings did differ as a function of judgement type. Again, participants acknowledged that undoing the crash is not enough to undo the final outcome. These results, in conjunction with results from Experiment 2, seem to favour an explanation of counterfactual alternative generation in line with the primacy constraint hypothesis (Kahneman & Miller, 1986; Miller & Gunasegaram, 1990; Wells et al., 1987). This hypothesis states that people tend to focus on the first element of a chain when attributing causality, given that the first element is not constrained by later events. Participants in Experiment 3 have consistently attributed higher ratings to what it is considered the first element in a causal chain in both the causal and counterfactual task.

General Discussion

The experiments presented in this paper were designed with the objective of testing the Judgement Dissociation Theory (Mandel, 2003b) and to determine whether causal selection and counterfactual reasoning 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, it was shown that the restricted coding is not the only factor responsible for this effect: The match in the instructions is another contributing factor, also observed in the first experiment. In Experiment 3, the hypothesis of task interference was discarded as responsible for the ratings trends observed in the previous experiments. The first two experiments also evidenced the ambiguity inherent in counterfactual manipulations, and how ambiguity is a function of the tasks demands.

This set of experiments has repeatedly failed to replicate the dissociation between causal and counterfactual reasoning and consequently this study does not offer support for JDT. Whereas the predicted match between causal and counterfactual reasoning did not emerge as clearly as expected, there are some elements that offer insight not only regarding the systematicity of the effect but also on the shortcomings of JDT.

Experiments 1 (medium level), 2 and 3 show a similar pattern of responses and ratings. 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 phrased in the most specific level of description (i.e., Experiment 1), where crime life was given the lowest ratings and chosen by the smallest proportion of people. In other words, only when the task referred to the specific way in which the main character died, was the causal dependence on crime life not considered. This is also reflected in the mean base rate of death that participants provided in the probability ratings. In comparison to the change in probability from this base rate compared to the action of any of the other elements, the probability was not as large as in Mandel's experiments.

This calls into question the a priori distinction between what constitutes a sufficient element for an effect 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 presented show that what is considered a preventor (and a sufficient cause) does depend on the particular description of the event, the instructions and the type of task demand. That is, these notions cannot be defined independently of the context (Hart & Honoré, 1959), but only in very limited settings.

There was, however, a consistent dissociation across the experiments regarding the importance of the direct cause (crash) in causal and counterfactual judgements. Whereas in the first experiment, crash was considered both the cause and the best way to undo the outcome only in the high level condition, in Experiments 2 and 3 crash was not considered as effective in the counterfactual as in the causal task. However, this tendency is actually complementary to the results observed in terms of attributions to crime, and is consistent with the primacy constraint hypothesis. The first element of the chain was considered both the most causal and the most effective in undoing the outcome, whereas the last element of the chain is totally determined by its predecessors and cannot effectively undo the outcome.

The variation is partially explained by the task demands. The instructions were matched, and that factor alone reduced the extent of the dissociation (Experiment 2). This result is related to the overlap between the questions about causes, explanations and counterfactuals, as pointed out by Hilton et al. (1990; McClure et al., 2007). Participants could have interpreted the questions about the factors causally responsible for the outcome as a more general question about an explanation of why things turned out to be that way. This is then the contrast between answering how and why questions about a causal situation. Although Mandel recognizes this important distinction in his later work (Mandel, 2005, 2011), he does not relate it to the link between causal and counterfactual reasoning, and claims that dissociation is functional, not the result of task demands and pragmatics.

Furthermore, it is possible that people represented the scenario in different ways. In fact, there are at least two alternative representations that consequently imply slightly different counterfactual dependencies. These are presented in Figure 1.

The first row of Figure 1 represents the scenario as a typical case of pre-emption, where counterfactuals dependencies do not allow singling out a cause, and both poison and crash are independently caused. The second row incorporates a new arrow to represent the possible influence on both the elements after poison and its impact on the final outcome. In this case, there is not an obvious candidate for causal selection. In this case, the immediate cause of the death (crash) is represented as an effect itself of poison. Take for example the answer of one of the participants:

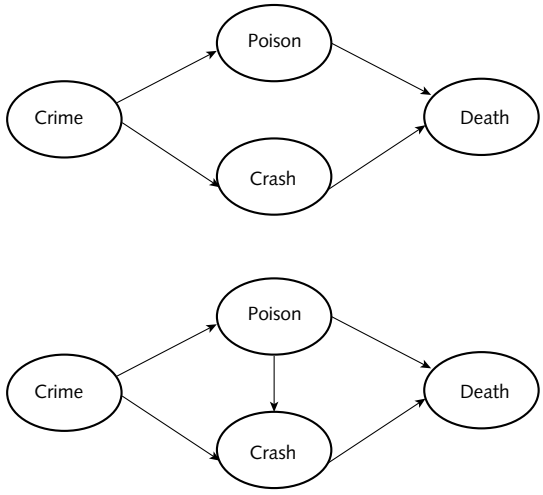

Figure 1. Alternative representation of the scenario used in the experiments

"If Mr Wallace had not drunk the poison, his reflexes would have been better, and he could have driven away from the van to his second appointment".

What these representations have in common is that counterfactual modification of any the nodes does not undo the final outcome unless the first one is modified. It then appears that participants considered this causal representation and recognized that interventions in any node were not sufficiently effective to undo the death of the protagonist. That is, in pre-emption scenarios, people cannot hold constant other elements in the causal network in order to assess the counterfactual impact of the interventions on it, particularly when it is possible that the two cause candidates are not independent from each other.

This also seems to suggest that cases of pre-emption are not suitable for testing theories of causal selection. It is very telling that cases of pre-emption are customarily used in discussions about philosophy of causation given that shared common intuitions about causality *fail* when analysing these cases.

It is still possible that the dissociation found by Mandel is a side product of the differential demand imposed by the causal and counterfactual tasks combined with the mismatch in the description in his causal and counterfactual tasks. Another possible explanation is that the general description of the outcome triggered a causal model that made available counterfactual options that were not available when causal factors involved in the death of the main character were requested (Sloman, 2005). Even in this case, the dissociation seems to be due more to a task demand than necessarily to a functional difference.

References

Ahn, W., Kalish, C., Medin, D. L., & Gelman, S. A. (1995). The Role of covariation versus mechanism information in causal attribution. *Cognition*, *54* (3), 290-352.

Friendly fire' hits Kurdish convoy. (2003, April). *BBC News*. Retrieved from http://news.bbc.co.uk/2/hi/middle_east/2921743.stm

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 (Eds.), *Rational models of cognition* (pp. 295-313). Oxford ; New York: Oxford University Press.

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

Einhorn, H. J. & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin Psychological Bulletin*, *99* (1), 3-19.

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

Hart, H. L. A. & Honoré, T. (1959). *Causation in the law*. Oxford: Clarendon Press.

Hilton, D. & Erb, H. P. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking and Reasoning*, *2*, 273-308.

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

Hoerl, C., McCormack, T., & Beck, S. R. (2011). *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology*. Oxford; New York: Oxford University Press.

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* (pp. 210 -210). Cambridge; New York: Cambridge University Press.

Kelley, H. (1983). Perceived causal structures. In J. M. F. Jaspars, F. D. Fincham, & M. Hewstone (Eds.), *Attribution theory and research: Conceptual, developmental, and social dimensions*. London; New York: Academic Press.

Lewis, D. (1973). *Counterfactuals*. Oxford: Basil Blackwell.

Lewis, D. (1987). *Philosophical papers volume II*. New York: Oxford University Press.

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. 132* (3), 419-434.

Mandel, D. (2005). Counterfactual and causal explanation: From early theoretical views to new frontiers. In D. R. Mandel, D. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking*. London; New York: Routledge.

Mandel, D. (2011). Mental simulation and the nexus between causal and counterfactual explanation. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology*. Oxford; New York: Oxford University Press.

Mandel, D. & Lehman, D. R. (1996). Counterfactual thinking and ascriptions of cause and preventability. *Journal of personality and social psychology*, *71* (3), 450-463.

McClure, J., Hilton, D., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European journal of social psychology*, *37* (5), 879.

Menzies, P. (2011). The role of counterfactual dependence on causal judgements. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.), *Understanding counterfactuals, understanding causation: Issues in philosophy and psychology*. Oxford; New York: Oxford University Press.

Miller, D. T. & Gunasegaram, S. (1990). Temporal order and the perceived mutability of events: Implications for blame assignment. *Journal of personality and social psychology*, *59* (6), 1111-1118.

N'gbala, A. & Branscombe, N. R. (1995). Mental simulation and causal attribution: When simulating an event does not affect fault assignment. *Journal of Experimental Social Psychology*, *31* (2), 139-162.

Pearl, J. (2000). Causation, action and counterfactuals. *Proceedings of the 6th conference on Theoretical aspects of rationality and knowledge*. The Netherlands: Morgan Kaufmann Publishers Inc.

Petrocelli, J. V., Percy, E. J., Sherman, S. J., & Tormala, Z. L. (2011). Counterfactual potency. *Journal of personality and social psychology*, *100* (1), 30-46.

Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford: Oxford University Press.

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

Spellman, B. A., Kincannon, A., & Stose, S. (2005). The relation between counterfactual and causal reasoning. In D. R. Mandel, D. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 41-60). London; New York: Routledge.

Teixeira, S. & Pacheco, X. (2001). *Borland Delphi 6*. Indianapolis Ind.: Sams Publishing.

Verschueren, N., Schaeken, W., & Schroyens, W. (2006). Necessity and sufficiency in abstract conditional reasoning. *The 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.

Wells, G., Taylor, B. R., & Turtle, J. W. (1987). The undoing of scenarios. *The Journal of Personality and Social Psychology*, *53* (3), 421-430.

Woodward, J. (2006). Sensitive and insensitive causation. *Philosophical Review Philosophical Review*, *115* (1), 1-50.