Conditions for Backtracking with Counterfactual Conditionals

Jung-Ho Han (JungHo_Han@Brown.Edu)[1]
William Jiménez-Leal (W.JimenezLeal@Uniandes.Edu.Co)[2]
Steven A. Sloman (Steven_Sloman@Brown.Edu)[1]
[1]Cognitive, Linguistic, and Psychological Sciences, Brown University, Box 1821
Providence, RI 02912 USA
[2] Departamento de Psicología, Universidad de los Andes.
Cra. 1 No. 18A-12, Edificio Franco, Bogotá, 17111. Colombia

Abstract

Counterfactual conditionals concern relations in other possible worlds. Most of these possible worlds refer to how a situation would have unfolded forward from a counterfactual assumption. In some cases, however, reasoning goes backward from the assumption, a phenomenon that is called backtracking. In the current study, we propose that people backtrack if and only if doing so will make a counterfactual claim true in the alternative world. We present evidence to support the proposal.

Keywords: counterfactual backtracking; causality; inference.

Introduction

Counterfactual conditionals are used in a variety of situations, from figures of speech ('if wishes were horses, beggars would ride') to causal inference ('if policy X had been implemented, millions of dollars could have been saved'). Recent psychological research has tried to clarify the link between counterfactuals and causal inference (Sloman & Pearl, 2013, for reviews), inspired by ideas from the causal modelling framework (Pearl, 2000). Briefly, the guiding hypothesis has been that counterfactuals are represented using a special kind of operator that consists of *intervening* on a variable in a causal model in order to infer its effects. Such interventions consist of locally modifying the actual value of the variable, while disconnecting from its causal ancestors. In this context, counterfactual reasoning about the implementation of policy X enables one to draw conclusions about the possible causal consequences of the policy, but does not give information about what other factors would have had to change for the policy to have been introduced.

Attention has focused on backtracking counterfactuals, a special type of counterfactual conditional whose antecedent allows inferring the value of upstream variables (Dehghani, Iliev, & Kaufmann, 2012; Rips, 2010; Rips & Edwards, 2013; Sloman & Lagnado, 2005). Consider, for example, the following conditional: "If the alarm had not gone off, it would have meant that I did not set it up correctly". In this case, the antecedent of the counterfactual is diagnostic of an earlier cause. While it is clear that this inference also depends on the appropriate causal representation of the world, it seems to fall outside the scope of the account proposed within the causal modelling framework (Sloman & Lagnado, 2005) because if the antecedent (the alarm clock not going off) were intervened on via the do operator, it would be rendered independent of its causes and hence non-diagnostic (therefore not implying that it had not been set up correctly). Some researchers have attempted to explain the meaning of this sort of counterfactual by either subscribing to a dual explanation, one for forward and one for backward counterfactuals (Dehghani et al., 2012; Rips, 2010; Rips & Edwards, 2013) or to an alternative unified model (Lucas & Kemp, 2012). In this paper we focus on some conditions that make backtracking possible when reasoning with non-backtracking counterfactuals.

How to Backtrack

Causal Bayes nets (Pearl, 2000) have been widely used to understand how people represent, and reason with, causal information. The power of this representation is derived from the use of the *do* operator, which allows reasoners to represent the effects of actions on a causal structure, and thus to make not only observational but also interventional inferences. The *do* operator sets the value of a variable (*do*(X=x)) which allows inference of the effects of X. The intervention is assumed to cut off the variable from its normal causes, thus rendering it non-diagnostic of those causes. Consider the case of a transitive causal relationship from A to B and then to C. Intervening on B produces a model where C is the effect of B (represented by the arrow from B to C), but the intervention on B provides no information about the state of A (represented by the grey line from A to B).

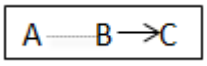

Figure 1: Transitive causal relationship.

Under certain conditions, people exhibit an *undoing* effect (non-diagnosticity of the intervened-on variable) and reason according to the logic of intervention (Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). Counterfactual conditionals can thus be conceived as an inference from an imagined intervention, where the antecedent is the variable intervened on, and the consequent is the effect read off from the causal model.

Rips (2010) has shown that the *do* operator does not apply in other cases of counterfactual reasoning. In his experiments, participants answered counterfactual questions about hypothetical mechanical devices, questions that

directly queried the state of upstream variables (e.g. "if component C were not operating, would component A be operating?"). Rips found that people often inferred the state of parent variables *contra* what is predicted by the interventionist approach; backtracking was common. The effect occurred with a selection of causal structures and depended on question wording, was less likely when relations were probabilistic (Rips & Edwards, 2013), and varies with the presentation order of the questions (Gerstenberg, Bechlivanidis, & Lagnado, 2013).

Two alternative theories have been proposed to explain how people backtrack with counterfactuals: minimal network theory (Dehghani et al., 2012; Rips, 2010; Rips & Edwards, 2013) and the double modifiable structural model (DMSM, Lucas & Kemp, 2012). Both accounts are based on the Bayes net formalism and use similar tools to explain counterfactual reasoning. Minimal network theory (Hiddleston, 2005), claims that when reasoning counterfactually, the changes introduced to the representation are minimal in that they respect the causal laws that govern the system. The idea is to keep the counterfactual model as similar as possible to the actual model; it must have as few edge-breaks and as many intact variables as possible (Rips, 2010). In consequence, to evaluate the changes introduced by the antecedent of a counterfactual, the causal connections that feed into the variable whose value has been changed need not be broken and thus backtracking is possible. This flexibility, however, makes the theory unsuitable for the case of reasoning about interventions (but see Dehghani et al., 2012).

DMSM proposes that reasoners hold an augmented twin network, a copy of the causal representation, which allows them to reason both from intervention and from observation. In the case of intervention, the model is equivalent to the use of the *do* operator. For observations, however, the augmented model includes a counterfactual representation of the exogenous variables that determine the value of the variables in the system. This captures the fact that the counterfactual world might turn out to be different from the real world even in the absence of interventions. DMSM includes a free parameter to represent the degree of mutability of the counterfactual model, which allows the model to offer good fits to published data on counterfactual backtracking (Lucas & Kemp, 2012).

These two theories build on the Bayes nets framework to allow for the possibility of backtracking. However, a full understanding of counterfactuals requires, from the point of view of the Minimal Network theory, two types of causal representation, one for intervening (Pearl, 2000) and one for backtracking, based on the alternative Minimal Network. On the other hand, DMSM requires a free parameter, whose psychological equivalent would be some sort of similarity weighting of possible worlds (Lewis, 1973).

While explaining how counterfactual backtracking takes place is certainly a key issue, an alternative approach is to determine *why* backtracking occurs. Most of the time, the introduction of a counterfactual supposition calls for changes in an asymmetric fashion. What if I had not gone to college? I probably would not have met your mother and I would be a lumberjack, etc. The consequences of the counterfactual supposition normally unfold into the possible future, and only in rare cases require backtracking. We believe that those cases that call for backtracking are tied to the need for explanation. The experimental setup of studies that have looked into backtracking are revealing in this respect: They explicitly ask about the state of an upstream variable given the counterfactually assumed antecedent ("if component C were not operating, would component A be operating?"). While it is likely that reasoners would normally evaluate only the downstream consequences of the counterfactual supposition, the experimental demands draw attention to information that might otherwise been ignored. Thus, by focusing attention on a set of previous causal factors, people put reasoning at the service of explaining why something could have come to be the case. This is closely related to similar ideas posited by Rips (2010) and Dehghani et al. (2012) about how the introduction of hypothetical beliefs involve adjustments to maintain consistency with prior knowledge.

In other words, the *explanation of* a counterfactual supposition might call for re-assessment of events that are causally upstream relative to the reference point introduced by the counterfactual antecedent. Alternatively, *reasoning from* the assumption of a counterfactual supposition (e.g. intervention) that calls for evaluation of events causally downstream that unfold from the counterfactual assumption. In this paper we evaluate two cases that require backtracking with conditional counterfactuals. The first case refers to conditionals whose antecedent and consequent are semantically independent but causally linked (Experiments 1 and 2). The second case refers to conditionals that express a causal link between antecedent and consequent, but where the effect requires the presence of an additional causal factor (Experiment 3). We hypothesize that reasoners backtrack if and only if they have to make a counterfactual conditional true.

Consider the following conditional statements, offered after learning that John has attended a birthday party.

If John weren't drinking alcohol, then he wouldn't have brought a gift. [1a]

If John weren't drinking alcohol, then he wouldn't act wildly. [1b]

Consequents of conditionals (1a) and (1b) are linked with their antecedents in two different ways. While the antecedent of conditional (1b) is causally responsible for the consequent, that is not the case for conditional (1a). Conditional (1a) only makes sense if a common cause, C, of both antecedent and consequent is assumed. Figure 1 graphically represents the underlying causal system for these statements.

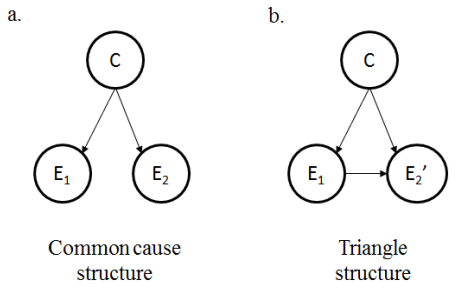

Figure 2: Causal structures tested in Experiment 1.

Note that the common cause, C, is not explicitly articulated by the speaker when uttering the counterfactual (1a); it is only assumed. We believe that backtracking occurs because interlocuters need to understand the mechanism that explains why the conditional is true (Figure 2a). By the same logic, backtracking is not necessary if the antecedent of a conditional is causally sufficient to infer its consequent (Figure 2b). According to our hypothesis, backtracking should occur for conditional (1a), but not for (1b) in a context in which the counterfactual is offered as a true statement. In Experiment 1, participants were asked to determine whether a speaker who states a counterfactual conditional, either (1a) or (1b), is also asking the listener to assume the presence of the common cause (e.g., John is at the party). In our scenarios, we presented participants with information that would explain the inferential connection between antecedent and consequent, namely the common cause.

In summary, in this paper we present the results of three experiments designed to test the hypothesis that backtracking occurs if and only if doing so allows reasoners to explain the truth of a counterfactual conditional. Experiment 1 uses a common cause structure to test the hypothesis with real life scenarios, whereas Experiment 2 does the same with abstract materials. Experiment 3 uses an alternative causal system, a common effect structure.

Experiment 1a

In this experiment we used the structures depicted in Figure 2 to construct conditional counterfactuals that linked the two effects in a conversational setting.

Method

Participants One-hundred-forty-five U.S. residents were recruited via Amazon's Mechanical Turk (AMT). Participation was restricted to workers in the United States.

Materials Seven scenarios were used to instantiate the causal structures shown in Figure 2. For example, the *common cause* structure was instantiated so that the two effects of a common cause (i.e. E_1 and E_2) took on the role of an antecedent and a consequent, respectively (Figure 2a). To implement the *triangle* structure, E_2 was replaced with a different consequent, E_2' (i.e. *E_2 prime*). E_2' was causally dependent on E_1 in addition to C (Figure 2b). Fourteen counterfactual conditionals were constructed in this way.

Participants were first informed of the factual state of the common cause and were asked whether it was being assumed when the conditional statement was spoken by someone in conversation. In the *common cause* conditions, for example, participants read a counterfactual conditional in which Abby tells Bonnie, "if John weren't drinking alcohol (~E_1), then he wouldn't have brought a gift (~E_2)," following the factual information about Joe's presence at a party (C). Participants then answered whether Abby was asking Bonnie to assume that John is NOT at a party (~C) when Abby told Bonnie the counterfactual conditional. Similarly, in the *triangle* condition, participants read a counterfactual conditional in which Bonnie tells Abby, "if John weren't drinking alcohol (~E_1), then he wouldn't act wildly (~E_2')," following the same factual information about C. Participants again answered whether Bonnie was asking Abby to assume that John is NOT at a party (~C) at the time when the counterfactual conditional was uttered. Participants used the mouse to choose "yes" or "no" for each question. We expect participants to answer "yes" in the *common cause* condition and "no" in the *triangle* condition.

Design Two causal structures (*common cause* vs. *triangle*) and seven scenarios were manipulated within-participants. Question items pertaining to the two structures were paired for each scenario and they were presented consecutively. Both the order of causal structures (common cause versus triangle) and the order of presentation of the scenarios (a single random order or its reverse) were counterbalanced across participants.

Procedure The first screen on the computer briefly explained the task to participants. It also clarified that they would be compensated only upon successful completion of the survey. The second screen asked an attention-check question to allow only participants who paid attention to instructions. The consent form was signed electronically on the third screen. Counterfactual questions were then presented over the next 14 screens. After participants completed the task, another attention-check question was presented. The data from individuals who failed the attention check were excluded from analyses. This left us with 128 participants who completed the experiment.

Results

Initial analyses revealed that responses were unaffected by the order in which causal structures and scenarios were presented. The results are thus collapsed over these factors.

Proportions of "yes" responses were first transformed using an arcsine square root transformation to increase the normality of the distribution. The t-test revealed that participants were more likely to respond "yes" to the *common cause* items (M = 39.01, SD = 19.27) than the *triangle* items (M = 29.78, SD = 21.65), $t(127) = 5.29$, p

< .001. That is, for the common cause structure participants were inclined to agree, across all scenarios, that the hypothetical speaker of a conversation was asking her listener to assume a change in the state of the common cause.

Experiment 1b

The results from Experiment 1a could be alternatively attributed to a stronger association between the common cause (C) and the consequent of the conditional for the *common cause* structure (E_2; see Figure 2a) compared with that of the *triangle* structure (E_2'; see Figure 2b). Putting it differently, participants have not shown backtracking for scenarios with the *triangle* structure because the common cause was 'remote' in those structures. If this is the case, it is necessary to measure the strength of association between the common cause and the consequent. In Experiment 1b, we directly asked participants the conditional probability of the common cause given the counterfactual state of the consequent for both structures. Our main hypothesis is supported should the current study reveals that the judged probability of the consequent in the *triangle* condition is higher than or equal to the *common cause* condition.

Method

Participants One-hundred-twenty-three U.S. residents were recruited via AMT.

Materials The same causal structures and scenarios from Experiment 1a were used to construct counterfactual conditionals. In the current experiment however participants were informed of the factual states of C and E_1. They were then asked to judge the probability of the counterfactual state of C given the counterfactual state of E_2 or E_2'. For instance, in the *common cause* condition, participants first read, "Candice is pregnant (C)" and "Candice is buying baby furniture (E_1)." Participants then provided the probability of Candice being pregnant if she were not buying baby furniture ($\sim E_2$). In the *triangle* condition, $\sim E_2$ was replaced by $\sim E_2$'. That is, participants provided the probability of Candice being pregnant if she were not gaining weight. The scale was from 0 to 100.

Design and Procedure The design and procedure were similar to those of Experiment 1a with changes to implement randomization of the items and probability judgment as a dependent variable.

Results

Data from 109 participants were analyzed as a result of excluding those who failed to complete the experiment. Probability judgments about the counterfactual state of C were higher in the *triangle* condition (M = 49.03, SD = 22.24) than in the *common cause* condition (M = 43.89, SD = 23.87), $t(108) = -5.17$, $p < .001$, supporting the conclusions from Experiment 1a. That is, the results were not due to the weak causal strength between C and E_2' in the *triangle* condition.

Experiment 2

Experiment 2 aims to replicate the finding of Experiment 1 with more abstract materials that do not lend themselves to content-dependent alternative explanations. In addition, it tests the effect with a between-participants design to see whether the effect depends on a direct comparison between the different kinds of counterfactuals.

Method

Participants and Design One of the two conditions (*common cause vs. triangle*) was randomly assigned to 73 Brown University undergraduates from 5 different psychology classes as in-class exercises. 37 participants were assigned to the *common cause* condition and 36 to the *triangle* condition.

Materials and Procedure We used materials from Rips and Edwards (2013) to frame the causal structures shown in Figure 2. Participants were given a sheet of paper with a description of a hypothetical device whose components operated the way they were graphically represented in Figure 2. For example, the description for the common cause condition stated:

> Professor McNutt of the Department of Engineering has designed a device called a glux. The glux has only three components, labeled A, B, and C. The device works in the following way:
> - Component A's operating causes components B to operate.
> - Component A's operating causes component C to operate.

For each device, the factual states of all components were described as 'currently not operating' Participants then judged the counterfactual state of A given the following counterfactual conditional: If B were operating, then C would be operating. The causal graphs were also provided in addition to the written description**.** Participants answered by circling "yes" or "no."

Results

The percentage of participants who answered "yes" and "no" differed significantly in the common cause condition, χ^2 (1, N = 37) = 4.57, p < .05 . That is, participants were more likely to respond "yes"(67.57%) in the common cause condition. However, the differences between "yes" and "no" were only marginally significant in the triangle condition, p = .096, (though the proportion of "no" (63.9%) was higher than "yes")

Experiment 3

Experiments 1 and 2 present evidence in favor of our hypothesis, but they do so only for a particular kind of causal structure. In order to further generalize our results,

we examined common effect structures in Experiment 3. To reiterate our hypothesis, we believe that people will backtrack only if they need to make the conditional true. In the case of common effect structures (see Figure 3), backtracking with counterfactuals antecedents about the state of an effect (*E* in Fig 3) can only happen when another causal factor acts *in conjunction* with the cause presented in the conditional consequent to generate the effect (Figure 3a). People do so because the second causal factor (*En* in Fig 3a) is a necessary condition for the effect to occur. On the other hand, people would not need to backtrack when an alternative causal factor (*aC* in Fig 3b) is presented, as it is an unnecessary piece of information for the truth of a counterfactual about *E* and *C*. In this case the occurrence of *C* (in Fig 3b) is sufficient for the truth of the conditional. In Experiment 3, we asked participants to consider whether they should assume an enabler (*En*) or an alternative cause (*aC*), when reasoning about counterfactual conditionals linking effect and cause. Our prediction is that backtracking for a conditional counterfactual in this case will only occur for conjunctive common effect structures.

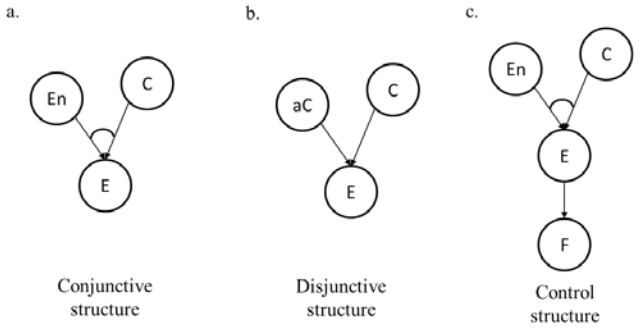

Figure 3: Causal structures used in Experiment 3. *En* is an enabler that operates in conjunction with cause *C* to produce E. *aC* is an alternative (disjunctive) cause.

Method

Participants One-hundred-twenty-three U.S. residents were recruited via AMT.

Materials Eleven scenarios were used to frame the underlying causal structures that are shown in Figure 3. For example, the *conjunctive* structure was framed such that E and C took on the role of an antecedent and a consequent, respectively (Figure 3a). The *disjunctive* structure was also framed like the *conjunctive* structure except that an enabler (En) was replaced with an alternative cause (aC) (Figure 3b). 22 counterfactual conditionals were constructed as a result. Questions were presented in a conversation format involving two imaginary characters. Depending on the causal structure they were given, participants judged the likelihood of En or aC being assumed by an imaginary listener when the conditional statement was spoken by someone. For example, in the *conjunctive* condition, participants read a dialog in which Abby tells Bonnie, "if I were buying the toy (E), then it would mean that my son's birthday was approaching (C)." Participants then judged the likelihood of whether Abby was asking Bonnie to assume that the toy was available (En), whereas in the *disjunctive* condition, participants judged the likelihood of whether Abby was asking Bonnie to assume that it is Christmas time (aC). Participants responded using 5-point Likert scales (1=Definitely not; 3=I don't know; 5=Definitely yes).

Control items were identical to the *conjunctive* condition except that the counterfactual conditionals were constructed with E, an antecedent, and F, a consequent (Figure 3c). For example, Abby tells Bonnie, "if I had bought the toy (E), then my son would have been very happy (F)." Again, participants judged the likelihood of whether Abby was asking Bonnie to assume that the toy was available (En).

Design Two causal structures (*conjunctive* vs. *disjunctive*) and eleven scenarios were manipulated as within-subject variables. Items pertaining to the two structures were paired for each scenario and they were presented consecutively without interruption. The pairs were randomized. Additionally, a between-participant factor was used to counterbalance the order in which the causal structures were paired, *conjunctive/disjunctive* or *disjunctive/conjunctive*. Participants were randomly assigned to one of the two counterbalancing conditions.

Control items were given to all participants.

Procedure The procedure was similar to that of Experiment 1a with changes to implement 5-point Likert scales.

Results

Ninety-seven participants successfully completed the experiment. Initial analyses revealed that responses were unaffected by the order in which the items were paired. The results are thus collapsed over this factor.

Participants judged that an enabler (En) in the *conjunctive* condition (M = 3.45, SD = .54) was more likely to be assumed than an alternative cause (aC) in the *disjunctive* condition (M = 1.95, SD = .55), $t(96) = 21.53$, $p < .001$.

In the control condition, participants judged an enabler (En) less likely (M = 2.92, SD = .56) than in the *conjunctive* condition (M = 3.45, SD = .54), $t(96) = 10.71$, $p < .001$.

Discussion

In this paper we have presented a novel variety of conditional counterfactual and presented evidence supporting the hypothesis that people backtrack if and only if they need to make the conditional true, when the conditional is offered in a conversational setting. We have tested our hypothesis using two causal structures, common cause and common effect. The use of a conversational context is a step toward examining counterfactuals in more realistic settings in which what is relevant is made clear.

Our hypothesis is agnostic about the best explanation of how people backtrack. However, it is not clear how it can be compatible with Minimal Network theory, since from that perspective either backward and forward inferences are equally likely (Edwards and Rips, 2013), Implementation by

DMSM (Lucas & Kemp, 2012) would be more straightforward. Our hypothesis does, however, imply a particular way of conceiving the origin of backtrackers.

Both Rips (2010, Edwards & Rips, 2013) and Deghani et al. (2010) present the issue of backtracking in relation to inference and explanation. However, they do it in slightly different ways. Deghani et al (2010) propose that backtracking is used to explain why the antecedent of the counterfactual is plausible. People backtrack because of "the speakers' desire to find a causal explanation for the hypothesized truth of the antecedent" (p. 64), and "explanations are likely to be implicitly involved in our evaluation of forward counterfactuals" (p.65). In contrast, Rips and Edwards (2013) suggest that backtracking occurs to explain why something was not the case, based on a similar idea by Sobel (2004). Their second experiment specifically asks people to explain, in the context of reasoning about a mechanism, why a component did not work: "One natural way to interpret counterfactual conditionals is to attempt to explain their antecedents" (Rips & Edwards, 2013, p. 24).

Our thesis is a slight, but important, departure in the interpretation of explanation and backtracking offered so far. We believe people backtrack to make sense of the proposed truth of the counterfactual conditional, not only of its antecedent. Backtracking, in general, occurs when one wants to explain why something might have been the case. However, backtracking does not occur haphazardly but respects the causal structure of the situation. In this sense, backtracking is a special case of explanation against a backdrop of stable conditions. We have shown that this set of stable conditions that allows backtracking can be delimited depending on the causal model that is built to represent a situation. Considered in this light, backtracking counterfactuals can be considered a case of causal belief revision determined by the structure of the situation (Sloman & Walsh, 2008; see Degahni et al. 2010, for a similar point).

Consider the following variation of a famous example (Adams, 1970):

If Oswald hadn't shot Kennedy, then Kennedy would have lived longer. [2]

If Oswald hadn't shot Kennedy, then someone else would have. [3]

While (2) requires reading off the values of an intervention performed on a local model, (3) asks us to roll "back history as we know it, and rerun it under different conditions" (Pearl, 2011, p.31). According to our thesis, only the second counterfactual requires backtracking because its truth value cannot be determined unless further assumptions about the preceding causes are made (e.g. public anger shared by other shooters). The situations considered by Rips (2010; Rips & Edwards, 2013) and Deghani et al (2012) are a special case of a more general pattern of explanation based on prior knowledge that occurs in contexts as varied as those offered in a conversation.

Acknowledgments

We thank Lance Rips for helpful discussion. W JiménezLeal was funded by the FAPA grant at Universidad de los Andes.

References

Adams, E. (1970). Subjunctive and indicative conditionals. *Foundations of Language, 6*(1), 89-94.

Dehghani, M., Iliev, R., & Kaufmann, S. (2012). Causal Explanation and Fact Mutability in Counterfactual Reasoning. *Mind & Language, 27*(1), 55-85. doi: 10.1111/j.1468-0017.2011.01435.x

Gerstenberg, T., Bechlivanidis, C., & Lagnado, D. A. (2013). Back on track: Backtracking in counterfactual reasoning. In M. P. Knauff, M., N. Sebanz & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Hiddleston, E. (2005). A Causal Theory of Counterfactuals. *Noûs, 39*(4), 632-657.

Hoerl, C., McCormack, T., & Beck, S. R. (2011). *Understanding counterfactuals, understanding causation: issues in philosophy and psychology*. Oxford: OUP.

Lewis, D. K. (1973). *Counterfactuals*. Cambridge: Cambridge University Press.

Lucas, A., & Kemp, C. (2012). A unified theory of counterfactual reasoning. In N. Miyake, D. Peebles & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 707-712). Japan.

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

Pearl, J. (2011). The algorithmization of counterfactuals. *Annals of Mathematics and Artificial Intelligence*, 61, 1, 29-39.

Rips, L. J. (2010). Two Causal Theories of Counterfactual Conditionals. *Cognitive Science, 34*(2), 175-221. doi: 10.1111/j.1551-6709.2009.01080.x

Rips, L. J., & Edwards, B. J. (2013). Inference and explanation in counterfactual reasoning. *Cognitive Science, 37*(6), 1107-1135. doi: 10.1111/cogs.12024

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

Sloman, S. A., & Pearl, J (2013) Rumelhart Prize Special Issue of Cognitive Science Honoring Judea Pearl Edited by Steven Sloman and Judea Pearl Vol 37 (6), 969–976.

Sloman, S. A., & Walsh, C. (2008). Updating Beliefs With Causal Models: Violations of Screening Off. In G. H. Bower, M. A. Gluck, J. R. Anderson & S. M. Kosslyn (Eds.), *Memory and mind: a festschrift for Gordon H. Bower*. New York: Lawrence Erlbaum Associates.

Sobel, D. M. (2004). Exploring the coherence of young children's explanatory abilities: Evidence from generating counterfactuals. *British Journal of Developmental Psychology, 22*(1), 37-58.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing Versus Doing: Two Modes of Accessing Causal Knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 216-227.