Challenges of Adding Causation to Richer Event Descriptions

Rei Ikuta*, William F. Styler IV+, Mariah Hamang*, Tim O'Gorman*, and Martha Palmer*
Department of Linguistics, University of Colorado at Boulder
*{rei.ikuta, mariah.hamang, ogormant, martha.palmer}@colorado.edu,
+will@savethevowels.org

Abstract

The goal of this study is to create guidelines for annotating cause-effect relations as part of the Richer Event Description schema. We present the challenges faced using the definition of causation in terms of counterfactual dependence and propose new guidelines for cause-effect annotation using an alternative definition which treats causation as an intrinsic relation between events. To support the use of such an intrinsic definition, we examine the theoretical problems that the counterfactual definition faces, show how the intrinsic definition solves those problems, and explain how the intrinsic definition adheres to psychological reality, at least for our annotation purposes, better than the counterfactual definition. We then evaluate the new guidelines by presenting results obtained from pilot annotations of ten documents, showing that an inter-annotator agreement (F1-score) of 0.5753 was achieved. The results provide a benchmark for future studies concerning cause-effect annotation in the RED schema.

1 Introduction: The RED schema and cause-effect relation

Richer Event Description (Styler et al., 2014a) is an annotation schema which is developed "as a synthesis of the THYME-TimeML guidelines[1], the Stanford Event coreference guidelines and the Carnegie Mellon University Event coreference guidelines." In other words, it combines Coreference (Pradhan et al., 2007; Lee et al. , 2012) and THYME Temporal Relations annotation (Styler et al. (2014b)) to provide a thorough representation of entities (including events) and their relations, including temporal relations. An overview of the annotation process, which shows how coreference and temporal annotations are combined, is described in the following section.

The RED schema therefore attempts to annotate cause-effect relations, which are annotated in neither Coreference nor THYME (Styler et al., 2014b). There is a synergy between annotating both causal and temporal relations, since causes necessarily precede their effects.

Other characteristics of the cause-effect annotation in RED are that it allows annotators to make inferences without relying on explicit connectives or verbs of causation, that it is not domain specific, and that it allows the relation to cross one (but not more than one) sentence boundary.

1.1 The annotation process

The process of RED annotation is divided into two passes: the first in which entities (including events) are annotated, and the second in which relations between those entities are annotated.

In the first pass, annotators identify three types of entities: events (an occurence with a definitive temporal duration), temporal expressions such as *August 2013*, and other entities that have an existence as opposed to an occurrence (e.g., proper nouns, objects, and pronouns). Specific properties of each event are also annotated in this pass (e.g., its relation to the document creation time, whether it is an actual event or a hypothetical event, etc.).

After the annotations in the first pass have been adjudicated, annotators mark temporal, cause-effect, and coreference relations between the entities identified in the first pass. Temporal relations (e.g., *before*, *overlaps*, *contains*) are annotated between two events or between an event and a TIMEX3, cause-effect relations are annotated between two events, and coreference relations are

[1]The THYME annotation schema also includes coreference annotation.

Proceedings of the 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation, pages 12–20,
Baltimore, Maryland, USA, June 22-27, 2014.

annotated between two entities (e.g., *President John F. Kennedy ... he*) or two events (*an earthquake ... the quake*). Coreference relations include part-whole and set-member relations, as well as identical relations in which two entities share a referent.

As a result of combining Coreference and THYME, different coreference and temporal relations between an event pair can be covered by a single relation in RED. For example, a part-whole relation between events annotated in Coreference (e.g., *an incision* and *a surgury*) is a subset of temporal "contains" relation in RED.

Therefore, the goal of RED is to combine Coreference and THYME annotation, while finding overlaps between the two and also introducing cause-effect annotation to achieve a richer representation of entities, events, and their relations.

1.2 Overview of the following sections

In the following sections, we present the challenges faced during our first pilot annotation and why we decided to change the definition of causation, from a counterfactual one to an intrinsic one. To support the use of the intrinsic definition, we examine the theoretical problems that the counterfactual definition faces, show how the intrinsic definition solves those problems, and explain how the intrinsic definition adheres to psychological reality, at least for our annotation purposes, better than the counterfactual definition. We then propose new guidelines based on the intrinsic definition and evaluate them by presenting results obtained from our second pilot annotations of ten documents, showing that an inter-annotator agreement (F1-score) of 0.5753 was achieved.

2 Challenges of cause-effect annotation using the counterfactual definition

The pilot annotations were done by three annotators who are native speakers of English and are experienced in linguistic annotation, on English proxy reports (i.e., approximations of intelligence agency reports) written by Garland, et al. (2013).

Our original guidelines were based on the counterfactual definition of causation, as defined below. Early on in the annotation process, the cause-effect annotation was halted and removed from the RED schema because there were a number of cases in which events matched our guidelines for the cause-effect relation but did not match our intuitions about the relation.

2.1 Counterfactual definition of causation

In the original guidelines for cause-effect relations, we defined causation as follows:

- "X caused Y" means if X had not occurred, Y would not have occurred.

This definition of causation in terms of counterfactual dependence (as philosophers call it) has been the most popular definition of causation in the field of philosophy for the past forty years since David Lewis's possible world model (Lewis, 1973) and remain influential in contemporary studies such as the structural model (Pearl, 2000; Halpern and Pearl, 2005).

Using this definition, one annotator marked two causal relations between the two event pairs in the following sentence[2]:

> (1) PYONGYANG INSISTS IT WILL **ALLOW** FULL IAEA **INSPECTIONS** ONLY WHEN A SIGNIFICANT PORTION OF THE **PROJECT** AS **DEFINED** IN THE 1994 ACCORD IS COMPLETED.
>
> Annotations:
>
> **ALLOW** causes **INSPECTIONS**
>
> **DEFINED** causes **PROJECT**

These annotations are done perfectly in line with the guidelines[3], since there would be no *inspections* if there were no *allowing*, and there would be no *projects* if there were no *defining* (of the project). Furthermore, one could argue that the *1994 accord* causes Pyongyang to *insist*, since if there had been no such accord, Pyongyang would not have been able to insist anything pertaining to it, although the annotators did refrain from creating such a causal annotation.

However, the relation between these event pairs does not match our intuition about what causation is. For example, the *allowing* should be considered as a precondition for the *inspections*, and not

[2]Another annotator who annotated the same text did not mark any causal relations in this sentence.

[3]The annotation guidelines allow future events to be in causal relations, although the counterfactual definition only deals with past events, and for quoted speech, narrators are assumed to be reliable. Thus, future events can participate in a causal relation if the narrator is certain about the relation. If the relation is presented to be likely or hypothetical instead of being actual, annotators can mark such modalities also.

the cause. Furthermore, the guideline creates too many event pairs that are potentially in a cause-effect relation (such as the *accord* and the *insisting*), contributing to confusion among annotators.

A similar issue can be seen in the following sentence, in which *the internet* should be considered as a possible precondition of *funding*, and not the cause:

> (2) THE WORKSHOP WILL STUDY THE USE OF THE INTERNET TO PROMOTE TERRORISM AND THE **INTERNET'S** ROLE IN FACILITATING MONEY TRANSACTIONS AND **FUNDING** TERRORIST GROUPS.
>
> Annotation:
>
> **INTERNET'S** causes **FUNDING**

Therefore, we concluded that the counterfactual definition of causation is not optimal for our annotation guidelines, and that we need an alternative definition of causation which does not rely on annotators to consider a possible world in which the cause does not occur.

Such an alternative definition, which we call the intrinsic definition, has been argued for by Menzies (1996; 1999; 2014). Such a definition treats causation as an intrinsic relation between events, meaning that it is "a local relation depending on the intrinsic properties of the events and what goes on between them, and nothing else" (Menzies, 2014).

Drawing on Menzies idea, we propose the following definition of causation which is being used in our new guidelines for cause-effect annotation:

- "X caused Y" means Y was inevitable given X.

With this definition, annotators would not have to consider any possible worlds in which an event did not occur in order to annotate cause-effect relations, and only have to focus on whether Y necessarily follows X, according to the context and their encyclopedic knowledge of the world.

In order to support our use of such a definition, we also present the challenges that the counterfactual definition faces in terms of theory and psychological reality in the following sections, and show how the intrinsic definition solves those problems.

3 Theoretical challenge of the counterfactual definition

The two situations below illustrate theoretical challenges which are faced by the counterfactual definition but not by the intrinsic definition.

3.1 Multiple causes

- There are three events (1, 2 and 3), and three individuals (A, B, and C). Events 1 and 2 occur at the same time, and event 3 follows events 1 and 2.
- In event 1, A shoots C in the head.
- In event 2, B shoots C in the heart.
- In event 3, C dies.
- Then, an autopsy reveals that each of the shots C received (one in the head, shot by A, and the other in the heart, shot by B) was sufficient by itself to kill C.

In the above situation (a modified version of the example in Lagnado et al. (2013)), the counterfactual definition would falsely predict that both events 1 and 2 are not the causes of event 3, since even if event 1 did not occur, event 3 would have occurred because of event 2, and if event 2 did not occur, event 3 would have occurred because of event 1.

Acknowledging this problem, Halpern and Pearl (2005) retain the counterfactual notion and extend their causal model by stating that counterfactual dependence should be evaluated relative to certain contingencies. According to this definition, the counterfactual dependence of event 1 to event 3 should be evaluated relative to a contingency in which event 2 does not occur. The obvious problem that this extended model faces is the difficulty of finding a principled way to decide which contingencies are allowed. Although Halpern and Pearl (2005) do offer a complex set of conditions that are aimed at capturing the intuition that one should only invoke contingencies "that do not interfere with active causal processes," the question of which contingencies are allowed is non-trivial and is the subject of ongoing debate (Halpern and Hitchcock, 2010; Hiddleston, 2005; Hopkins and Pearl, 2003; Lagnado et al., 2013).

This situation, however, is easily handled by the intrinsic definition, since event 3 (the death of C) is

inevitable given event 1 (A shooting C in the head) regardless of other events, and event 3 is inevitable given event 2 (B shooting C in the heart) regardless of other events, according to what we know about the results of the autopsy. Thus the intrinsic definition correctly predicts that both events 1 and 2 are equally the causes of event 3.

3.2 Oxygen, lightning, and wildfire

- There are three events (1, 2 and 3). Event 1 is a state encompassing events 2 and 3, and event 3 follows event 2.
- In event 1, oxygen exists.
- In event 2, a lightning strikes a tree in a forest.
- In event 3, a wildfire starts in the forest.

In this situation described by Halpern and Hitchcock (2013), event 1 (the existence of oxygen) would be predicted as being one of the causes of event 3 (wildfire), since if oxygen did not exist, a wildfire would not start. However, they argue that human intuition would treat only event 2, and not event 1, as a cause of event 3.

To counter this problem, Halpern and Hitchcock (2013) again extend the counterfactual model, stating that potential causes are graded according to the *normality* of their *witnesses* (a witness is a world in which a potential cause is the actual cause of an outcome). In this extended model, the world in which oxygen exists is more *normal* than the world in which lightning strikes a particular tree. Therefore, the lightning, being less normal, "receives a higher causal grading." In their causal model, a static ranking of the witnesses are given before the processing (i.e., causal inference) starts, and thus it is possible to compute which witness receives a higher causal grading.

Unlike the extended counterfactual definition, the intrinsic definition does not assume a given ranking of the world, and thus it is especially useful when applied to annotation tasks. For example, annotators would identify a causal relation between the oxygenation and the wildfire in the following sentence:

> (3) The **oxygenation** of the atmosphere accompanied by a lightning **strike** triggered the first **wildfire** in Earth's history.

But not in the following:

> (4) The first **wildfire** in Earth's history was caused by a lightning **strike** in the Proterozoic, an era noted for the evolution of multicellular organisms, glaciations, and the **oxygenation** of the atmosphere.

Even though the two events (*oxygenation* and *wildfire*) described in the above sentences refer to the same events in the world, the annotators can choose whether to note a causal link between them depending on the inevitability implied by the text. In sentence (2), it is suggested that the *wildfire* was inevitable given the *oxygenation* and the *strike*, thus both of the events would be annotated as the cause, while sentence (3) does not imply such a causal relation. This would effectively let the annotators avoid marking cause-effect relations between births and deaths in texts such as obituaries and medical reports. Such varying interpretations of texts are not possible with the original counterfactual definition, or with Halpern and Hitchcocks extended counterfactual model (2013) which assumes a given ranking of witnesses which is available to the writer but not to the annotator.

4 Challenge of the counterfactual definition in terms of psychological reality

In addition to the theoretical problem that the counterfactual definition faces, experiments done by White (2006) have shown that counterfactual dependence is not used as preferred evidence for making causal inference when subjects are passively (i.e., without the ability to intervene) exposed to a scenario in which there are a number of events affecting one another.

In one of the experiments, subjects are presented with scenarios concerning two game reserves, in each of which live five species, who may or may not prey on each other. For each reserve, there are five statements corresponding to five consecutive seasons, and each statement describes whether the population of each of the species has changed in that season. Based on the statements, the subjects must decide whether a change in the population of one species causes changes in that of the others. The subjects are instructed that if the population of X changed and that of Y did not in a given season, they are supposed to conclude that X does not prey on Y, because if it did, the populations of both X and Y would have changed.

In other words, the subjects are explicitly told to rely on counterfactual dependence as evidence for making causal inference. The five statements provided enough counterfactually dependent relations for the subjects to reach one correct answer.

However, the results of the experiment show that only 5 out of 36 subjects made correct judgments on the predator-prey (cause-effect) relations in both reserves, and the success rates were below optimum and not far above chance. Instead, the answers by the subjects showed that they were more likely to rely on the temporal order of events as the evidence for the causal relations (i.e., "the population of X changed in season 1 and that of Y changed in season 2, thus X must be the predator of Y"), although they were instructed to rely on counterfactual dependence within the same season instead.

White (2006) carried out three additional experiments, one in which he changed the order of the seasons, another in which subjects were told that the seasons were in random order and that the temporal order is irrelevant to the answer, and the last in which the scenario was changed to a situation where the levels of five chemicals in a blood stream affect each other. The subjects' answers exhibited more reliance on counterfactual dependence in the experiment where they were told that temporal order is irrelevant, but the other experiments showed similar results with the first experiment.

Thus, White (2006) concludes that there is a preference for basing causal inference on domain-specific causal knowledge (i.e., "the population change in season 1 must be causally related to the change in season 2, according to what we know about ecosystems") over counterfactual dependence, when such knowledge is available for use and when subjects are passively exposed[4] to a complex scenario in which there are a number of events affecting one another.

These results support our motivation to avoid using the counterfactual definition, since annotators are passively exposed to text without the ability to intervene, texts to be annotated are complex systems in which a number of events may or may not affect each other, and it is usually the case that domain-specific causal knowledge is available. The use of an intrinsic definition for cause-effect annotation, on the other hand, is in line with the results of these experiments, since annotators would not have to consider any possible worlds where some event does not occur, and only have to focus on whether Y necessarily follows X, according to the context and their encyclopedic knowledge of the world.

[4]It has been claimed that subjects perform better in making causal inferences on complex structures when they are actively exposed to (i.e., have the ability to intervene with) the structures (Lagnado and Sloman, 2004; Sloman and Lagnado, 2005; Steyvers et al., 2003).

5 The new guidelines

Given the challenges faced by the counterfactual definition and the advantages of the intrinsic definition presented above, we developed new guidelines for cause-effect annotation which instruct annotators as follows:

- In our schema, we annotate "X CAUSES Y" if, according to the writer, the particular EVENT Y was inevitable given the particular EVENT X.

We then utilized the counterfactual definition as the definition of precondition relations as follows:

- We annotate "X PRECONDITIONS Y" if, according to the writer, had the particular EVENT X not happened, the particular EVENT Y would not have happened.

The reason we kept the counterfactual definition in our guidelines as a definition of a precondition relation is that the relation defined by counterfactual dependence still gives us information about the temporal relation between events; if we know that Y would not have happened if X had not happened, we also know that X started before Y.

6 The second pilot annotation

Using the new guidelines, ten proxy reports were each annotated by two annotators. One of them was among the two annotators who participated in our first pilot annotation, and the other, who is also a native speaker of English experienced in linguistic annotation, was trained using the old guidelines but only started annotating in the RED schema after the cause-effect annotation was halted, and thus had not actually annotated cause-effect relations until the second pilot. The following sections present the inter-annotator agreement of cause and precondition annotations done in the ten reports and the analysis of specific examples where the annotators disagreed.

6.1 Inter-annotator agreement

This section presents the inter-annotator agreement (IAA) obtained from the second pilot annotation, and analyzes the annotations to examine the sources of disagreement between the annotators. Perhaps the most important thing to note before discussing the specific numbers and examples is that this pilot annotation did not include the adjudication stage between the first pass where entities including events and temporal expressions are identified, and the second pass where the relations between those entities are marked (see Section 1.1 for the specifics of the annotation process). Therefore, many of the disagreements in the causation and precondition annotations involve disagreements in the first pass.

A total of 114 relations (50 causation and 64 precondition relations) were created by the two annotators. Among them, 24 exhibited perfect match between the annotators, while 18 exhibited partial match (meaning that they agreed on whether the relation was causation/precondition, but disagreed on other aspects of the relation, such as the modality and temporal relation[5]) . Among the 114 relations, 72 relations showed disagreements, but 69 of them involved disagreements in the first pass. Upon analysis, we judged 41 of those 69 disagreements as being avoidable by introducing the adjudication stage between the two passes, and 28 as having the potential of surviving adjudication, meaning that even if the adjudication were properly done, the same parts of the text may still cause similar disagreements. Only 3 among the 72 disagreements occurred purely in the second pass, meaning that the annotators completely agreed on what the entities involved in the 3 relations should be, but disagreed on the relation.

Thus, the results give us four types of IAA (best-case, realistic, worst-case, and extra-strict), shown in Table 1 as F1-scores.

The best-case IAA assumes that all disagreements involving disagreements in the first pass will not show up as issues in the second pass, and only takes into account the 3 disagreements that occurred purely in the second pass.

	F1-score
Best-case	0.9333
Realistic	0.5753
Worst-case	0.3684
Extra-strict	0.2105

Table 1: Inter-annotator agreement for the second pilot annotation

The realistic IAA takes into account the 28 disagreements involving disagreements in the first pass that have the potential of surviving adjudication.

The worst-case IAA assumes that all disagreements in the first pass survive adjudication.

Finally, the extra-strict IAA allows relations to be judged as agreeing only when the two annotations completely match, including the modality and the temporal relations marked together with causation/precondition.

6.2 Evaluation of the inter-annotator agreement

This section compares the IAA presented above with results shown in a previous study by Styler et al. (2014b) which deals with temporal relation annotations in the clinical domain. In their study, Styler et al (2014b) reported results from annotations done on a subset of the THYME colon cancer corpus, which includes clinical notes and pathology reports for 35 patients diagnosed with colon cancer for a total of 107 documents. Two graduate or undergraduate students in the Department of Linguistics at the University of Colorado annotated each text. For the annotation guidelines, they used the THYME-TimeML guidelines which are also used within the RED guidelines for temporal relation annotation. Unlike the annotations in this current study, the temporal relation annotations on the THYME corpus were done after the identification of events and temporal expressions were adjudicated (the THYME-TimeML schema does not identify entities that are not events or temporal expressions). Therefore, the IAA they presented (Table 2) are not affected by the disagreements at the level of event identification.

The figure for "participants only" shows the IAA concerning cases in which the annotators

[5] As well as marking the modality (whether the relation is stated as being actual, likely or hypothetical) and the temporal relation (whether the cause ends before the effect starts or cause overlaps with the effect), annotators have a choice of marking a relation as "difficult" when they are not sure of their annotation. This difficulty marking was not considered when judging whether the two annotators agreed completely or not. In other words, even if one annotator marked a relation as difficult and the other did not, the annotation would be considered as showing complete agreement as long as other properties of the annotation matched.

	F1-score
Participants only	0.5012
Participants and relation	0.4506
”Contains” relaion	0.5630

Table 2: Inter-annotator agreement presented in Styler et al. (2014b)

agreed that there is some sort of a temporal relation between the two participants, but did not necessarily agree on which temporal relation (*before*, *overlap*, *contains*, etc.) holds between them. The figure for ”participants and relation” shows the agreement on both the participants and the type of the temporal relation. The third figure is the IAA for the temporal relation ”contains,” which exhibited the highest IAA among all the temporal relations.

These figures are significantly higher than the results reported for the 2012 i2b2 challenge (Sun et al., 2013), in which the F1-score for ”participants only” IAA was 0.39.

The realistic IAA of 0.5753 obtained in this current study is not far-off from the figures by Styler et al. (2014b), which shows that causation and precondition annotations using the new guidelines are indeed feasible.

6.3 Examples of disagreements

Below, we present examples of different types of disagreements observed in the annotations. The annotations are represented in the form of ”EVENT relation-relation EVENT.” The first half of the relation indicates the temporal relation annotated between the events, and the latter half shows whether there was a causation or a precondition relation between the events. For example, ”P before-cause Q” indicates that event P happened before and caused event Q.

6.3.1 Disagreement in the 1st pass: avoidable by adjudication

(5) A **BUDGET** WAS **ALLOCATED** FOR THE BARRIER TO BE **EQUIPPED** WITH ELECTRONIC DETENTION EQUIPMENT.

Annotations by annotators X and Y:
X: **ALLOCATED** before-preconditions **EQUIPPED**

Y: **BUDGET** before-preconditions **EQUIPPED**

In (5), annotator X marked *allocated* as an event while not marking *budget* as an event, and Y annotated *budget* as an event and did not mark *allocated* as an event. If the adjudication was correctly done, only marking *allocated* as an event and not *budget*, it is likely that Y would have annotated the same way as X.

6.3.2 Disagreement in the 1st pass: not avoidable by adjudication

(6) CRITICS STATE THAT WITH ACCESS TO PLUTONIUM AVAILABLE FROM ROGUE STATES TERRORISTS COULD **CONSULT** THE DETAILED DOCUMENTS AND **BUILD** AN ATOMIC BOMB.

Y: **CONSULT** before-preconditions **BUILD**

X: No relations identified

In (6), the annotators did disagree on whether the two events *consult* and *build* happen after or overlap with the document creation time (DocTime). X annotated those two events as overlapping the DocTime, while Y annotated them as after the DocTime. The annotators agreed that those two events were hypothetical events. Although such a disagreement about the temporal property of the events may have caused the disagreements about whether there should be a precondition relation, it is likely that X would have missed what Y had found even if there had been adjudication.

6.3.3 Disagreements in the 2nd pass

(7) THE SMH AND JENNINGS WERE THEN **SUED** OVER 3 ARTICLES **PUBLISHED** IN THE LEAD-UP TO THE 000000 OLYMPICS.

X: No relations identified

Y: **PUBLISHED** before-preconditions **SUED**

(8) HEAD OF A TAJIK GOVERNMENT AGENCY THAT FIGHTS DRUG TRAFFICKING AVAZ YULDACHEV STATED THAT HEROIN USERS ARE **ILL** AND **NEED** TREATMENT.

X: **ILL** overlap-cause **NEED**

Y: No relations identified

(7) and (8) above show cases in which one annotator missed the relation that the other annotator identified, even though both annotators completely agreed on the property of the entities involved in the relation.

7 Conclusion

In this paper, we have presented the challenges that the counterfactual definition of causation faces in terms of its application to annotation guidelines, theory, and psychological reality. We have shown that the intrinsic definition better suits our purpose of annotation, and proposed new guidelines for annotating cause-effect relations using such a definition. The new guidelines were evaluated using results obtained from a pilot annotation of ten documents. An inter-annotator agreement (F1-score) of 0.5753 was obtained. We are currently in the process of training four additional annotators with the new guidelines, and future studies concerning cause-effect annotation in the RED schema can assess their performances by using results presented in this paper as a benchmark.

Acknowledgments

The project described was supported by DARPA FA-8750-13-2-0045, subaward 560215 (via LDC) DEFT: Deep Exploration and Filtering of Text and NIH: 1 R01 LM010090-01A1, THYME (via Harvard). The content is solely the responsibility of the authors and does not necessarily represent the official views of DARPA or NIH.

References

Garland, J., Fore, D., Strassel, S., and Grimes, S. 2013. *DEFT Phase 1 Narrative Text Source Data R1 LDC2013E19*. Web download file. Philadelphia: Linguistic Data Consortium

Halpern, J. Y., and Hitchcock, C. 2010. Actual causation and the art of modeling. In R. Dechter, H. Geffner,and J. Y. Halpern, eds., *Heuristics, probability and causality: A Tribute to Judea Pearl*. (pp. 383–406). London: College Publications.

Halpern, J. Y., and Hitchcock, C. 2013. Compact Representations of Extended Causal Models. *Cognitive Science*, 37:986–1010.

Halpern, J. Y., and Pearl, J. 2005. Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*, 56(4):843–887.

Hiddleston, E. 2005. A causal theory of counterfactuals. *Nous*, 39(4):632–657.

Hopkins, M., and Pearl, J. 2003. Clarifying the usage of structural models for commonsense causal reasoning. In P. Doherty, J. McCarthy, M. Williams, eds., *Proceedings of the AAAI Spring Symposium on Logical Formalization of Commonsense Rea-soning*. (pp. 83–89). Menlo Park, CA: AAAI Press.

Knobe, J., and Fraser, B. 2008. Causal judgment and moral judgment: Two experiments. In W. Sinnott-Armstrong, eds., *Moral psychology, Volume 2: The cognitive science of morality*. (pp. 441–447). Cambridge, MA: MIT Press.

Lagnado, D. A., Gerstenburg, T., and Zultan, R. 2013. Causal Responsibility and Counterfactuals. *Cognitive Science* 37:1036–1073.

Lagnado, D. A., and Sloman, S. 2004. The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 30:856–876.

Lee, H., Recasens, M., Chang, A., Surdeanu, M., and Jurafsky, D. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Jeju Island, 489-500.

Lewis, D. 1973. Causation. *The Journal of Philosophy*, 70(17):556–567.

Menzies, P. 1996. Probabilistic Causation and the Preemption Problem. *Mind*, 105:85–117.

Menzies, P. 1999. Intrinsic versus Extrinsic Conceptions of Causation. In H. Sankey, ed., *Causation and Laws of Nature*, Kluwer Academic Publishers, pp. 313–29.

Menzies, P. 2014. Counterfactual Theories of Causation. In E. N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*. Retrieved from `http://plato.stanford.edu/archives/spr2014/entries/causation-counterfactual/`

Pearl, J. 2000. *Causality*. Cambridge: Cambridge University Press.

Pradhan, S. Ramshaw, L., Weischedel, R., MacBride, J., and Micciulla, L. 2007. Unrestricted Coreference: Indentifying Entities and Events in OntoNotes. In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC)*, September 17-19.

Sloman, S., and Lagnado, D. A. 2005. Do we "do"? *Cognitive Science*, 29:5–39.

Steyvers, M., Tenenbaum, J. T., Wagenmakers, E. J., and Blum, B. 2003. Inferring causal networks from observations and interventions. em Cognitive Science, 27:453–489.

Styler, W., Crooks, K., O'Gorman, T., and Hamang, M. 2014a. *Richer Event Description (RED) Annotation Guidelines.* Unpublished manuscript, University of Colorado at Boulder.

Styler, W. F., Bethard, S., Finan, S., Palmer, M., Pradhan, S., de Groen, P. C., Erickson, B., Miller, T., Lin, C., Savova, G., and Pustejovsky., J. 2014b. Temporal Annotation in the Clinical Domain, *Transactions of the Association of Computational Linguistics*, 2:143–154.

White, P. A. 2006. How well is causal structure inferred from cooccurrence information? *European Journal of Cognitive Psychology*, 18 (3):454–480.