Causal Premise Semantics

Abstract
The rise of causality and the attendant graph-theoretic modeling tools in the study of counterfactual reasoning has had resounding effects in many areas of cognitive science, but it has thus far not permeated the mainstream in linguistic theory to a comparable degree. In this paper I show that a version of the predominant framework for the formal semantic analysis of conditionals, Kratzer-style Premise Semantics, allows for a straightforward implementation of the crucial ideas and insights of Pearl-style Causal Networks. I spell out the details of such an implementation, focusing especially on the notions of intervention on a network and backtracking interpretations of counterfactuals.

1 Introduction
Conditional reasoning is a fundamental part of human life. As an object of research, this basic human capacity touches on a number of important questions surrounding uncertainty and chance, time and causation, metaphysics and epistemology. Not surprisingly, therefore, it has long been a focus of attention in linguistics, philosophy, psychology, and artificial intelligence. For the most part, each discipline approached the topic with its own theoretical questions and methodological tools, while there was little interdisciplinary interaction. It was only around the turn of the millennium that a confluence of ideas began to emerge which brought the need for cross-disciplinary exchange back into relief. Particularly conspicuous in this regard is the rise of causal (in)dependencies as an irreducible notion in its own right, amenable to theoretical modeling and empirical verification. The rise and spread of this perspective in turn owes much to the pioneering development and untiring advocacy of Judea Pearl.

These developments have implications for linguistic theory. A proper understanding of the meaning and use of conditionals has been one of the enduring goals of linguistic research for the last several decades. This work stands to benefit from the study of the underlying cognitive processes in neighboring disciplines. Conversely, too, since research on hypothetical and causal reasoning deals to a large
extent in linguistic data, linguistic theory has potentially much to contribute to the formulation of hypotheses and to the design and interpretation of empirical studies in other disciplines.

The predominant framework in the formal semantic analysis of modals and conditionals is the Premise Semantics introduced by Kratzer (1977, 1978, 1981a,b, 1991a,b).\(^1\) Kratzer’s work played a pivotal role in bringing then-recent methodological developments from philosophical logic to linguistics. The recent causal turn with its attendant graph-theoretic modeling tools has not had a comparable impact on the linguistic mainstream, its transformative effect in other areas of cognitive science notwithstanding. I believe that such parochialism is not beneficial for either side. Although it remains to be seen whether and to what extent Causal Networks and Premise Semantics may inform or enrich each other, nothing can come of it without some understanding of what the central tenets of one amount to in the context of the other.

This paper proposes an integrated approach which draws on both. The approach reflects a certain view of the relationship between them: Premise Semantics as a framework is encompassing enough to accommodate the crucial assumptions and insights behind the theory of Causal Networks. This claim may seem odd because Premise Semantics does not deal in probabilities, whereas Causal Networks remain tightly connected to the context of Bayesian reasoning in which they were first developed. But this latter association is not a necessary one. The notions of independence and intervention that networks are used to capture are not essentially tied to a probabilities. My goal is to show what a Premise Semantic model would have to look like in order to support these kinds of inferences. At the same time, in terms of the formal apparatus and its underlying ontological commitments I aim to stay close to the standard Kratzer-style framework. In this regard, the present proposal differs from related prior work on counterfactuals, including my own (Kaufmann, 2001a,b,c, 2004, 2005a, 2009), but also that of Kratzer (Kratzer, 1989, 2002) and Schulz (2007, 2011). I return to some of these differences below.

## 2 Preliminaries and basic assumptions

In linguistic semantics, conditionals are analyzed as modal expressions. For instance, (1b) is just a variant of (1a); the role of the ‘if’-clause is to specify (hypothetical) assumptions relative to which its truth is to be evaluated.

\[
\begin{align*}
1 & \quad \text{a. John must be at home.} \\
& \quad \text{Must(John home)}
\end{align*}
\]

\(^1\)Some of these papers were re-issued in updated form under the same titles in Kratzer (2012). Since the updates do not affect the arguments in this paper, I continue to cite the original versions.
b. If the lights are on, John must be at home. \( \text{Must}_{\text{lights on}}(John \text{ home}) \)

The same analysis applies to conditionals headed by other modals, such as 'may', 'can', 'might', etc. Thus the treatment of modal expressions in this framework forms the backdrop for the analysis of conditionals.\(^2\)

The clause in the scope of a modal operator (e.g., 'John be at home' in (1a,b)) is called its **prejacent**. Modalized sentences are about consequence or **consistency** of the prejacent relative to some body of information. Which relationship is involved is a question of **modal force**. In many languages including English, the force of a modal is specified by its lexical meaning (e.g., 'must' and 'may' express consequence and consistency, respectively). Which body of information is relevant is a question of **modal flavor**. Most modals are compatible with multiple modal flavors: For instance, (1a,b) can be interpreted **epistemically** (relative to what is known) or **deontically** (relative to norms or obligations).\(^3\) The modal flavor is typically left implicit, although it can be made explicit, for instance, by adding prefixes like those in (2). In semantic analysis, the modal flavor is treated as a contextual parameter.

\[(2)\]
\[
a. \quad \text{[In view of the available evidence,] John must be at home. [epistemic]}
\]
\[
b. \quad \text{[In view of the regulations,] John must be at home. [deontic]}
\]

The conditionals that this paper is about are variously called “counterfactual” and “subjunctive.” Neither term perfectly characterizes all and only the sentences in the relevant class, but I so not attempt to sort out the thorny taxonomical implications of this situation, both for reasons of space and because such a discussion would distract from the main concerns of this paper. Instead, for present purposes, suffice it to say that I take the relevant class, under either name, to include (3a,b) but not (3c).

\[(3)\]
\[
a. \quad \text{If this match were struck, it would light.}
\]
\[
b. \quad \text{If this match had been struck, it would have lit.}
\]
\[
c. \quad \text{If this match is struck, it will light.}
\]

\(^2\)The modal analysis is also applied to conditionals without a modal auxiliary, such as (i).

\[(i)\]
\[
\text{If the lights are on, John is at home.}
\]

In the interest of a unified analysis for all conditionals, such cases are assumed to involve a default modal operator akin to epistemic 'must' (see below). I have no occasion to rely on this assumption in this paper and do not discuss it further. See Gillies (2010); Khoo (2011) for recent discussion.

\(^3\)Linguistic theory distinguishes more modal flavors (see Portner, 2009, for an overview).
2.1 The auxiliary 'will'

What the sentences in (3) above have in common is that they are headed by the auxiliary verb 'will', in its Past form in (3a,b) and in the Present in (3c). In this subsection I briefly lay out a couple of basic assumptions about 'will' and 'would' which help locate the present proposal in the context of linguistic theory. They are not uncontroversial, but I stop short of offering a detailed linguistic argument. The reader is invited to consult the references cited for further background.

2.1.1 'Will/would' is a modal

'Will' and 'would' are inflectional forms of a common root whose meaning is essentially modal, not unlike 'may' and 'might' or 'shall' and 'should'. This implies that 'will' is not, or at least not primarily, a marker of Future tense. Rather, its forward-shifting effect on the reference time, where it arises, is due to a combination of factors including its Present tense, its modal lexical meaning, the aspectual properties of its prejacent, and interactions with other elements within the same sentence (such as temporal adverbs) and beyond (such as contextually given reference times). In support of this view, notice that 'will' does not necessarily involve future reference (witness the parenthesized addition in (4a)) and that even on its ordinary futurate use as in (4b) it adds an element of "conjecture" or "prediction" (see Kaufmann, 2005b, for details and references).

(4) a. John will be at home right now.
   b. John will be at home tomorrow.

This notion of conjecture or prediction is most relevant here. I assume that the modal root underlying both 'will' and 'would' expresses necessity relative to what is normal or stereotypical given certain facts or assumptions. There is room for variation as to what those facts or assumptions are. While the most prominent reading for (4) is epistemic (expressing what the speaker infers or is disposed to infer from certain pieces of information), 'will' can also be interpreted objectively. For instance, (3c) above need not express the speaker’s subjective assessment, but can instead convey a factual, objectively verifiable claim about a causal process “out there” in the world.

---

4The assumption that 'will/would' is a modal is common in linguistics (Abusch, 1997; Sarkar, 1998; Condoravdi, 2002; Copley, 2002; Kaufmann, 2005b), though not universally shared (Kissine, 2008; Groman and von Stechow, 2011).

5This description does not cover a number of unrelated uses of 'will/would', most prominently (i) as a main verb (as in 'She willed her heart to stop'); (ii) as a modal auxiliary with a volitive reading (as in 'If he will help us, we'll be grateful'); and (iii) in generic statements like 'Boys will be boys'.
This is the reading that this paper is primarily concerned with. It is worth re-emphasizing, before moving on, that the subjective/objective distinction just discussed is not treated as a lexical ambiguity in Kratzer-style semantics, but rather as arising from different settings of the contextual parameters affecting the interpretation. Consequently, the present analysis is semantic/pragmatic in that it is aimed at explicating how literal meaning and contextual parameters interact to deliver a particular kind of interpretation.

2.1.2 ‘Would’ marks counterfactualty

The second point to be made about ‘will/would’ concerns the role of their temporal morphology, specifically the Past tense of ‘would’. I treat this as an instance of so-called “fake Past” (Iatridou, 2000), the co-opting of Past or Perfect morphology to refer to states of affairs that are hypothetical or unreal, and not necessarily prior to the utterance time.6 There are competing ideas on how exactly this modal use of the Past tense comes about. It is fair to say that none of them has as-yet been fully developed into an account of all the relevant ingredients (including the difference between ‘would’ in (3a) and ‘would have’ in (3b)) and their interaction.7 Nor is it the goal of this paper to propose such an analysis. What I will say, by way of preview of the proposal to follow, is that I take ‘would’ in counterfactuals to signal that the crucial parameters of the interpretation are manipulated so as to “accommodate” or “make room for” the hypothesis expressed in the antecedent. What this means will become clear once the formal framework is spelled out in some more detail, and I will return to it later on in the paper. The notions of revision, intervention and explanation are crucially important in this connection.

2.2 Kratzer-style Premise Semantics

To facilitate the exposition, I start by assuming that English sentences are translated to logical forms in the standard language of propositional modal logic. A model for the interpretation of this language is a structure \( \langle W, V \rangle \), where \( W \) is a non-empty set (of “possible worlds”) and \( V \) is a function mapping sentences in the language

---

6See Jespersen (1924); James (1982); Fleischman (1989); Dahl (1997); Iatridou (2000); Kaufmann (2005a); Schulz (2012), among others. Across languages which have fake Past, counterfactual conditionals seem to be its prototypical environment, although it may also be encountered in other constructions, e.g., in the complement of ‘wish’ in English.

7In addition to the references in Fn. 6, see Arregui (2005b, 2007, 2009); Ippolito (2006), and references therein.
to subsets of $W$, subject to the condition that the logical connectives receive their usual interpretation:\footnote{8}{The backslash ‘\’ denotes set subtraction. Other truth-functional connectives can be defined in terms of these as usual.}

\begin{align}
(5) \quad & \text{a. } V(\overline{\varphi}) = W \setminus V(\varphi) \\
& \text{b. } V(\varphi \land \psi) = V(\varphi) \cap V(\psi)
\end{align}

I refer to sets of worlds as \textbf{propositions}, following standard usage in the linguistic literature. Thus sentences $\varphi$ and the propositions $V(\varphi)$ they denote are distinct kinds of entities. However, for convenience I will often write ‘$\varphi$’ instead of ‘$V(\varphi)$’. The notion of truth is defined for both propositions and sentences. A proposition $p \subseteq W$ is true at a world $w \in W$ if and only if $w \in p$, and true in the model if and only if it is true at all of its worlds. In both of these senses, a sentence $\varphi$ is true if and only if $V(\varphi)$ is true.

\subsection{2.2.1 Modality}

The next step is to extend the language by adding modal operators. First, I define the crucial logical properties and relations for their interpretation.

\textbf{Definition 1} \textit{(Consistency and consequence)}

Let $P$ be a set of propositions and $p$ a proposition.

\begin{enumerate}
\item $P$ is \textbf{consistent} iff there is a world at which all propositions in $P$ are true.
\item $p$ is a \textbf{consequence} of $P$ iff $p$ is true at all worlds at which all propositions in $P$ are true.
\item $p$ is \textbf{consistent} with $P$ iff $p$ is true at some world at which all propositions in $P$ are true.
\end{enumerate}

As discussed above, the meaning of modals is analyzed along two dimensions, modal force and modal flavor. As a first approximation (to be refined below), we can say that modal force is modeled as consequence and consistency. In introducing the parameter of modal flavor above, I referred to a “body of background assumptions” relative to which the consequence or consistency of the prejacent is evaluated. Formally, these background assumptions are derived via the interplay between two parameters: The first is the \textbf{modal base}, representing what is \textit{established} in the relevant sense (i.e., \textit{known} or firmly \textit{believed} for epistemic readings; \textit{true} for objective ones). The second is an \textbf{ordering source} representing what is \textit{preferred} in the relevant sense (i.e., \textit{likely} or \textit{normal} for stereotypical readings; \textit{required} for deontic ones; \textit{desired} for bouletic ones; etc.). In contrast to the invio-
lable content of the modal base, the ordering source contributes information which is defeasible and may be internally inconsistent (as may be the case, for instance, with wishes or laws).

In interpreting modals, the basic idea is to add propositions from the ordering source to the modal base while maintaining consistency. The sets of propositions generated in this way are the **premise sets**.

**Definition 2** (Kratzer premise sets)
Let $M, O$ be two sets of propositions. The set $\text{Prem}^K(M, O)$ of **Kratzer premise sets** contains all and only the consistent supersets of $M$ obtained by adding (zero or more) propositions from $O$.

If all possible ways of assembling premise sets inevitably lead to a premise set which entails a given proposition $p$, then $p$ is a **necessity** relative to $M$ and $O$. The notion of **possibility** is the dual of this.\(^9\)

**Definition 3** (Kratzer necessity and possibility)
Let $\Phi$ be a set of premise sets and $p$ a proposition.

a. $p$ is a **necessity** relative to $\Phi$ iff every premise set in $\Phi$ has a superset in $\Phi$ of which $p$ is a consequence.

b. $p$ is a **possibility** relative to $\Phi$ iff there is some premise set in $\Phi$ such that $p$ is consistent with all of its supersets in $\Phi$.

The formal implementation in Kratzer’s framework adds only one wrinkle to this: Since the contents of “what is known,” “what is the law,” and similar notions are themselves contingent, the modal base and the ordering source are modeled as functions mapping possible worlds to sets of propositions (rather than as sets of propositions directly). Kratzer calls functions of this type **conversational backgrounds**. It is customary in linguistics to use the variables $f$ and $g$ to refer to the modal base and ordering source, respectively. The interpretation of modalized sentences relative to $f, g, w$ is then defined as expected:

**Definition 4** (Kratzer interpretation of modals)
Let $f, g$ be conversational backgrounds and $w$ a possible world.

a. $\text{Must}(p)$ is true at $f, g, w$ iff $p$ is a necessity relative to $\text{Prem}^K(f(w), g(w))$.

b. $\text{May}(p)$ is true at $f, g, w$ iff $p$ is a possibility relative to $\text{Prem}^K(f(w), g(w))$.

'Will(p)' is similar to 'must(p)', modulo any difference in the modal parameters.

\(^9\)Further gradations and comparative notions of necessity and possibility are offered by Kratzer, but not relevant to my concerns in this paper.
This does not yet cover ‘would(p)’, which on the relevant use does not occur outside of conditionals, i.e., without an (either explicitly or implicitly given) antecedent. I turn to conditionals next.

### 2.2.2 Conditionals

As mentioned above, in Kratzer’s framework, ‘if’-clauses modify modal operators. More precisely, the role of ‘if(p)’ is to transform the relevant modal base $f$, creating a new modal base that I will refer to as $f[p]$. Kratzer derives $f[p]$ by simply adding $p$:

**Definition 5** (Kratzer update)

Let $f$ be a conversational background. The **Kratzer update** of $f$ with $p$, $f[p]$, is a conversational background such that for all possible worlds $w$, $f[p](w) := f(w) \cup \{p\}$.

Conditionals are interpreted by evaluating the modalized main clause relative to a modal base that is modified with the antecedent.

**Definition 6** (Kratzer interpretation of conditionals)

Modal $\text{Modal}_p(q)$ is true at $f, g, w$ iff Modal $\text{Modal}(q)$ is true at $f[p], g, w$.

Under an epistemic interpretation of $f$, the update corresponds to “hypothetically adding” the antecedent to speakers’ “stock of knowledge,” Ramsey’s (1929) famous paraphrase of the reasoning involved in evaluating indicative conditionals. But Ramsey explicitly reserved this operation for speakers who are uncertain about the truth of the antecedent. This is typically not the case for counterfactuals. If $p$ cannot be added to $f(w)$ consistently, some adjustments have to be made (see Stalnaker’s 1968 adaptation of the Ramsey Test to counterfactuals).

Kratzer avoids this potential problem by ensuring that inconsistency does not arise in the first place. To facilitate the discussion, some relevant terminology is fixed in Definition 7.

**Definition 7** (Properties of conversational backgrounds)

A conversational background $\beta$ is said to belong to the following classes iff the corresponding property holds for all worlds $w, u, v \in W$:

- **consistent**: $\beta(w)$ is consistent.
- **empty**: $\beta(w)$ is the empty set.
- **realistic**: all propositions in $\beta(w)$ are true at $w$.
- **unique**: if all propositions in $\beta(w)$ are true at $v$ and $u$, then $v = u$.
- **totally realistic**: $\beta$ is realistic and unique.
Kratzer assumes that counterfactuals are generally interpreted relative to the *empty* modal base; call it \( f_0 \). Since \( f_0(w) \) is the empty set of propositions irrespective of \( w \), the antecedent \( p \) can always be added consistently, as long as \( p \) itself is not a contradiction.

Now, if the modal base for the interpretation of a counterfactual invariably contributes the singleton set containing the antecedent, then no non-trivial information about the world of evaluation is preserved. Factual information about \( w \) does enter the interpretation, however, via the ordering source \( g \).\(^{11}\) Kratzer stipulates that for counterfactuals, \( g \) has to be totally realistic; i.e., \( g(w) \) must uniquely characterize the world \( w \) of evaluation. This ensures that when the antecedent is true at \( w \), the counterfactual is true if and only if its consequent is.\(^{12}\) Most of Kratzer’s work on counterfactuals revolves around the question what else should be required of the ordering source for the case that the antecedent is false at \( w \).

Admitting *all* propositions that are true at \( w \) would trivialize the semantics for counterfactuals. For let a “naïve” ordering source \( g^n \) be one whose value at \( w \) is the set of all propositions that are true at \( w \) – i.e., \( g^n(w) = \{ p \subseteq W \mid w \in p \} \). It can be shown that in this case, if the antecedent is false at \( w \) then the counterfactual is true at \( w \) if and only if its consequent logically follows from \( p \).\(^{13}\) Clearly this is not a sensible prediction: Counterfactuals do not require for their truth that their antecedent entail their consequent.

Kratzer (1981b) sought to remedy this problem by imposing further constraints on premise sets while maintaining crucial aspects of her overall approach, most notably the use of an empty modal base and a totally realistic ordering source. This combination remains problematic, however. It implies, roughly speaking, that the

\(^{10}\)Kratzer (1991b) points out the correspondence with modal logic: For each conversational background \( \beta \), there is a unique accessibility relation \( R_\beta \) defined as follows: \( wR_\beta v \) if \( v \in \bigcap \beta(w) \). The converse mapping is not unique: One and the same accessibility relation may correspond to multiple distinct conversational backgrounds.

Some of the properties in Definition 7 of conversational backgrounds correspond to well-known properties of accessibility relations in modal logic, with some variation in terminology: e.g., consistent \( \rightarrow \) serial; realistic \( \rightarrow \) reflexive.

\(^{11}\)The underlying idea can be traced back to Goodman’s (1947) proposal that counterfactuals assert that the consequent can be inferred from “the conjunction of the antecedent and other statements that truly describe relevant conditions.” This intuition underlies all flavors of Premise Semantics (Rescher, 1964; Veltman, 1976, 2005; Loewer, 1979; Lewis, 1981; Pollock, 1981; Kratzer, 1981b, 1989, 2002; Schulz, 2007, 2011, and references therein).

\(^{12}\)The condition of total realism corresponds to *strict centering* in the ordering semantics of Stalnaker (1968) and Lewis (1973). See Lewis (1981) for details.

\(^{13}\)Suppose \( p \) is false at \( w \) and the consequent does not follow logically from \( p \). Then there is a world \( v \in p \) at which the consequent is false. Notice that the proposition \( \{ w, v \} \) is in \( g^n(w) \) and consistent with \( p \). Hence there is a set \( \{ p, \{ w, v \} \} \in \text{Prem}^n(f[p](w), g^n(w)) \), all of whose consistent supersets imply the falsehood of the consequent. See Veltman (1976); Kratzer (1981b) for further details.
maximal premise sets are always maximally similar to the world of evaluation. But counterfactuals are not always interpreted in terms of maximal similarity. Consider for instance the following example after Kratzer (1989): In the zoo, an absent-minded helper forgot to lock the door to an enclosure which held two zebras and a giraffe. One of the zebras escaped. Against this context, consider the sentences in (6):

(6) a. If a different animal had escaped, it would have been a zebra.
   b. If a different animal had escaped, it might have been a giraffe.

Intuitively (6a) is false and (6b) is true; however, it is not clear how to account for this intuition in a framework that is built around similarity to the world of evaluation. The proposition that the escaped animal was a zebra is true and consistent with the antecedent, yet the judgments suggest that it is somehow barred from membership in the relevant premise sets: For otherwise the maximal premise sets would contain it and entail the consequent, hence the judgments would be reversed.

Kratzer’s (1989) response was an appeal to a relation of “lumping” between propositions. (The label is taken from a suggestion by Lewis, 1981.) While this is not the place for a detailed discussion of that proposal, I briefly mention three reasons for exploring an alternative route. First, Kratzer’s proposal relies on ontological commitments (to a hierarchy of “situations” or partial worlds, and to cross-world counterpart relations between these situations) which Kratzer claims are needed to give content to the lumping relation, but which I see no need to adopt. Second, there are open questions concerning the coherence of the formal implementation (see the debate in Kanazawa, Kaufmann, and Peters, 2005; Kratzer, 2002, 2005). Third, “lumping” as a concept seems amorphous and hard to grasp independently of its use in the semantics of counterfactuals. Although causality is not without its detractors, it does enjoy widely shared intuitive support, something lumping cannot claim for itself as far as I can tell.\textsuperscript{14}

Causal (in)dependencies have been appealed to twice before in the linguistic literature on counterfactuals. Kaufmann (2001a,b,c, 2004, 2005a, 2009) developed a semantics for indicative and counterfactual conditionals in which causal networks play an important role in correcting certain counterintuitive predictions of the Bayesian approach of Adams (1965, 1975) and others. But that analysis was couched in a probabilistic framework, whereas in this paper I am specifically interested in a Premise Semantic account. Kaufmann’s theory also stops short of offering an account of the “backtracking” reasoning I discuss in Section 4.1 below.

\textsuperscript{14}That said, it is conceivable that the role played here by causal relationships is one way in which lumping might come about. I will not speculate on this possibility here, but I briefly return to it at the end of the paper.
Schulz (2007, 2011) based a broadly Premise Semantic analysis (in the framework of Veltman, 2005) on causal networks. Schulz’s approach differs from mine in a number of respects, most notably the analysis of what I call backtracking resolutions of counterfactuals (see Section 4.1 below). Adjudicating between the two accounts would require a detailed comparative study of particular examples, which would be beyond the scope of the present paper and must wait for another occasion.

2.3 The formal framework

My causal account is couched in a Premise Semantic framework, but the formal implementation generalizes the Kratzer-style system just discussed in some respects. The difference in implementation concerns the way in which premise sets are derived and ranked. The Kratzer-style derivation in Definition 2 above operates on two sets of propositions \( f(w), g(w) \), provided by the modal base and ordering source at the world of evaluation, and calls for the addition of (zero or more) propositions from the latter to the former. Among all possible ways of doing so, the consistent ones are the premise sets. Thus for instance, if \( f(w) \) and \( g(w) \) are as in (7a), the set of premise set is (7b).

\[
\begin{align*}
\text{(7)} & \quad \text{a. } f(w) = \{p, q\}, \quad g(w) = \{r, s\} \\
& \quad \text{b. } \text{Prem}^K(f(w), g(w)) = \{\{p, q\}, \{p, q, r\}, \{p, q, s\}, \{p, q, r, s\}\}
\end{align*}
\]

This rule embodies two assumptions of Kratzer’s Premise Semantics: the only subset of \( f(w) \) that may occur in the premise sets is \( f(w) \) itself, whereas all subsets of \( g(w) \) are available to construct premise sets. Both of these assumptions are highly constraining. To see this, it is useful to switch to a generalization of the formal setup that makes the space of alternatives explicit.

Let us say that a **premise background** is a function mapping possible worlds to sets of sets of propositions. I use bold-faced variable names like \( f, g \) to refer to these objects. The similarity to Kratzer’s notation is intended, as I am particularly interested in premise backgrounds which stand in some systematic relationship to the modal bases and ordering sources discussed so far. Specifically, I say that a premise background \( f \) structures a given conversational background \( f \) if at all worlds \( v, f(v) \) is a set of subsets of \( f(v) \) containing \( f(v) \) itself. This is a non-trivial but still rather weak constraint. Importantly for present purposes, it ensures that \( f \) is necessarily (i.e., at all worlds) no more or less informative than \( f \).

---

15Schulz draws a distinction between “ontic” and “epistemic” interpretations of counterfactuals, but the underlying intuitions are quite different from the ones I propose below in terms of reasoning to the best explanation.

16The constraint also ensures that \( f(v) \) is upper-bounded, a property that is preserved when two premise backgrounds are compared along the lines discussed in this section. Premise backgrounds
To continue the simple illustration from above, let $f_{id}$, $g_\wp$ be two premise backgrounds defined as in (8). Their values at $w$ are given on the right.

(8) a. $f_{id}(v) := \{f(v)\}$ for all $v$. 
   $f_{id}(w) = \{\{p, q\}\}$

b. $g_\wp(v) := \wp(g(v))$ for all $v$.
   $g_\wp(w) = \{\emptyset, \{r\}, \{s\}, \{r, s\}\}$

Both $f_{id}$ and $g_\wp$ structure the corresponding conversational backgrounds. Consider now the Cartesian product of $f_{id}(w)$ and $g_\wp(w)$, i.e., the set of all pairs (of premise sets) drawn from the two premise backgrounds at $w$. In order to avoid clutter, I simplify the notation and write ‘$ab.cd$’ instead of ‘$((a, b), (c, d))$’, and ‘$ab.$’ instead of ‘$((a, b), (\emptyset))$’.

(9) $f_{id}(w) \times g_\wp(w) = \{pq., pq.r, pq.s, pq.rs\}$

Clearly each pair in (9) corresponds to a unique way of adding (zero or more) propositions from $g(w)$ to $f(w)$, i.e., to a premise set in Kratzer’s system.17 There is a straightforward way to evaluate modal claims relative to such a set of pairs. The underlying ideas are very simple: First, in terms of content, a pair of premise sets is identified with its union. In other words, its linear structure is ignored: What matters is which propositions it contains, not in which position those propositions occur. Thus for instance, $pq.r$ is consistent iff $\{p, q, r\}$ is, and $t$ is a consequence of (consistent with) $pq.r$ iff $t$ is a consequence of (consistent with) $\{p, q, r\}$.

Secondly, in determining the ranking of the premise sets for the interpretation of modals, the internal structure of the pairs is not ignored. Rather, in each pair the left position takes priority, while the propositions on the right only come into play for pairs that are tied in the left position. Formally, this means that the pairs are ordered lexicographically – that is, according to the following rule.

(10) $(X, Y) \leq (X', Y')$ iff (i) $X \subseteq X'$ and (ii) if $X = X'$ then $Y \subseteq Y'$.

For the pairs in (9) above, all of which contain $\{p, q\}$ in the left position, the lexicographic order breaks down when there is variation in both positions. This becomes evident when pairing $f_\wp(w)$ with $g_\wp(w)$. The sixteen pairs together with the resulting ranking are shown in Figure 1(b). For instance, while the premise sets $\{p, r, s\}$ and $\{p, q, r\}$ are incomparable in terms of set inclusion (and of the same cardinality), the corresponding pairs $p.rs$ and $pq.r$ are asymmetrically ranked with respect to each other by the lexicographic order.

---

17The inverse correspondence is not unique if there are propositions that are in both $f(w)$ and $g(w)$. Such overlap introduces redundancy in the premise set pairs, which however has no ill effect on the resulting truth conditions of modal sentences.
Figure 1: Hasse diagrams of the lexicographic orders obtained by pairing various subsets of \( \wp(\{p, q\}) \) with \( \wp(\{r, s\}) \).

For another example, consider a premise background \( f_c \) structuring \( f \), whose value at world \( w \) is neither the singleton set containing \( f(w) \) nor its powerset, but something in between:

\[(11) \quad f_c(w) := \{\emptyset, [p], [p, q]\}\]

The result of pairing this set with \( g_{\wp}(w) \) is shown in Figure 1(c). The label ‘c’ on \( f_c \) stands for “causal,” since the set in (11) has properties that will be important in the causal account of counterfactuals below.

The above examples illustrate the combination, point-wise at individual worlds, of a pair of sets (of sets of propositions) into a set of pairs. This procedure can be generalized by extending the notion of a premise set to include such pairs as well as longer sequences of premise sets. (In Definition 9, ‘\( \leq_1 \ast \leq_2 \)’ stands for the lexicographic order on the Cartesian product.)

**Definition 8 (Sequence structure)**

The set of sequence structures is recursively defined as follows:

a. If \( \Phi \) is a set of sets of propositions, then \( (\Phi, \subseteq) \) is a sequence structure.

b. If \( (\Phi_1, \leq_1), (\Phi_2, \leq_2) \) are sequence structures, then so is \( (\Phi_1, \leq_1) \ast (\Phi_2, \leq_2) \), defined as \( (\Phi_1 \times \Phi_2, \leq_1 \ast \leq_2) \).
In words, a sequence structure is a set whose members are sequences of (one or more) sets of propositions. If $\Phi$ is a basic sequence structure, it is ordered by set inclusion, which is a partial order. Since this property is preserved by the lexicographic product, all sequence structures are partially ordered. Furthermore, since both Cartesian products and lexicographic orders are associative, so is the operation of product formation on sequence structures, i.e., the equivalence in (12) holds for all sequence structures $X, Y, Z$:

\[(12) \quad X * (Y * Z) \equiv (X * Y) * Z\]

Definition 8 is general enough to allow for the combination of arbitrarily many premise backgrounds. That such flexibility is needed is suggested by recent appeals to go beyond the Kratzer-style two-parameter framework, for instance by combining primary with secondary ordering sources (von Fintel and Iatridou, 2008; Katz et al., to appear). In this paper, however, I only deal with sets of sequences of up to three premise sets.

In a slight but convenient abuse of notation, I use the same operator ‘*' to extend the notion of premise background to include functions from possible worlds to sequence structures of arbitrary length. If $f, g$ are two premise backgrounds, then $(f * g)(w) := f(w) * g(w)$ for all worlds $w$.

A sequence structure represents all ways of combining premise sets from the respective backgrounds; Definition 8 does not require them to be consistent. The sequences that are used in interpretation modal statements should of course be consistent, just like the premise sets in Kratzer’s system.

**Definition 9** (Premise structure)

Let $\langle \Phi, \leq \rangle$ be a sequence structure. The **premise structure** $\text{Prem}(\langle \Phi, \leq \rangle)$ is the pair $\langle \Phi', \leq' \rangle$, where $\Phi'$ is the set of consistent sequences in $\Phi$ and $\leq'$ is the restriction of $\leq$ to $\Phi'$.

Since sequence structures are partially ordered, so are the premise structures derived from them, as this property is preserved under restriction.

With these notions in place, the definitions of the relevant logical relations parallel those from Definition 3 above, with premise sequences in place of premise sets.

**Definition 10** (Necessity and possibility)

Let $\langle \Phi, \leq \rangle$ be a premise structure and $p$ a proposition.

- $p$ is a **necessity** relative to $\langle \Phi, \leq \rangle$ iff for every $X$ in $\Phi$ there is a $Y$ in $\Phi$ such that $X \leq Y$ and $p$ is a consequence of $Y$.  

b. \( p \) is a possibility relative to \( \langle \Phi, \leq \rangle \) for some \( X \) in \( \Phi \), \( p \) is consistent with every \( Y \in \Phi \) such that \( X \leq Y \).

The interpretation of modals in the new framework is also parallel to the earlier one in Definition 11. I maintain the differentiation between a modal base \( f \) and an ordering source \( g \). Both may themselves be complex, built up from more basic backgrounds, and both are combined into one for the purpose of deriving the premise structure, so the two-way split is arbitrary from a formal point of view. However, conceptually and in their role as representing the contextual parameters, the distinction is real and significant. Moreover, in Section 4.1 I will have occasion to rely on the distinction in formulating a constraint on premise sequences.

**Definition 11** (Interpretation of modals)

Let \( f, g \) be premise backgrounds and \( w \) a possible world.

a. Must(\( q \)) is true at \( f, g, w \) iff \( q \) is a necessity relative to \( \text{Prem}(\langle f \ast g \rangle(w)) \).

b. May(\( q \)) is true at \( f, g, w \) iff \( q \) is a possibility relative to \( \text{Prem}(\langle f \ast g \rangle(w)) \).

Finally, the interpretation of conditionals follows the general strategy of adding the proposition denoted by the antecedent to the modal base. More specifically, the antecedent proposition is paired with premise sets contributed by \( f(w) \). In analogy with the above discussion, this operation is an “update”; however, I do not mean to claim that the same operation is appropriate for modeling permanent belief change outside of the hypothetical reasoning involved in the interpretation of conditionals. As an explicit reminder of this caveat, I call the operation a hypothetical update.

**Definition 12** (Hypothetical update)

The result of hypothetically updating a premise background \( f \) with a proposition \( p \) is a premise background \( f[p] \) defined as follows, for all worlds \( w \): \( f[p](w) := \{\{p\}\} \ast f(w) \).

The set \( \{\{p\}\} \) in the definition is a simple set of premise sets – the single premise set containing the single proposition \( p \). The result at a world \( w \) is, informally speaking, the set of sequences obtained by prefixing \( p \) to all members of \( f(w) \). For instance, if the premise background \( f_{id} \) from (8a) is hypothetically updated with a proposition \( r \), then the result at \( w \) is (13a). For another example, the update of \( f_{c} \) from (11) results in (13b).

\[ (13) \quad \text{a.} \quad f_{id}[r](w) = \{\{r\}\} \ast f_{id}(w) = \{r, pq\} \]

\[ \text{A permanent change in belief might be better represented by actually adding the new proposition to all sets in } f(w), \text{ rather than pairing it with } f(w) \text{ in the manner of Definition 12.} \]
With the hypothetical update in the background, the interpretation of conditionals is defined in analogy with the earlier operation in Definition 6 above.

**Definition 13 (Interpretation of conditionals)**

Modal\(p(q)\) is true at \(f, g, w\) iff Modal\(q\) is true at \(f[p], g, w\).

Recall from that above that both \(f_{id}\) and \(f_c\) structure the Kratzerian modal base \(f\) (with \(f(w) = \{p, q\}\)), although \(f_{id}\) does so in a trivial fashion. This is where we can locate the difference between indicative and counterfactual conditionals. I assume that all conditionals generally presuppose that the hypothetical update with their antecedent yields a consistent premise background (i.e., one which holds at least one consistent premise sequence at all worlds of evaluation). If the antecedent \(r\) is consistent with \(f(w)\), then \(f_{id}[r]\) succeeds. What is special about counterfactuals is that they call for a modal background which structures the modal base non-trivially, i.e., whose value at each world \(w\) contains some proper subset of \(f(w)\) in addition to \(f(w)\) itself. As long as some element of \(f(w)\) is consistent with the antecedent at all worlds \(w\), the semantic requirement is met. Now, how exactly the modal base is structured is a matter of pragmatics. In this paper I am concerned with causal interpretations. I turn next to the properties of modal backgrounds that make them suitable for this reading.

### 3 Causal Premise Semantics

In the statistical jargon underlying the theory of causal networks, the most basic part of the model is an outcome – simply put, a way an experiment or observation may turn out. On the formal semantic side, the analog of an outcome is a possible world. Sets of outcomes are called events, corresponding to propositions in the possible-worlds framework. In statistics, a random variable induces a partition of the space of possible outcomes into a mutually exclusive and jointly exhaustive set of events, intuitively corresponding to a particular “question” one might ask about the process under investigation. The corresponding notion in possible-worlds semantics can be thought of as a question denotation.\(^{19}\) These correspondences are listed in Table 1.

Basically, a causal network consists of two parts, dedicated to the representa-\(^{19}\)

\(^{19}\)While some semantic theories model the denotations of questions as partitions of the logical space (notably Groenendijk and Stokhof, 1984), others do not. The assumption that questions denote partitions implies a certain view on what constitutes a (semantic) answer. This is not the place to go into details on that topic, however.
The first component is a directed acyclic graph (DAG) whose vertices are variables and whose edges indicate direct causal influence. A DAG is a binary relation whose transitive closure is a strict partial order (i.e., irreflexive). The second part is a representation of the dependencies between the variables which is sensitive to the structure of the DAG: Each variable depends only on the values of its immediate parents in the graph. How exactly this is encoded is a matter of the underlying assumptions about the variables and the nature of the dependency: The values of the parents may determine either the value of the variable in question (in case of deterministic causation), or the probability distribution over its values (for non-deterministic causation). Accordingly, the dependencies may be encoded as systems of equations, as conditional probabilities, or in some other suitable fashion. What is important is that once the values of the parents are given, no further information about other non-descendants of the variable in question can affect its value or probability.

I do not assume that causality is deterministic, for two reasons: First, determinism is a special case of non-determinism, but not vice versa, and there is (and possibly can be) no proof that the world is deterministic. Secondly, regardless of what physicists tell us about the way the world works, we talk as if certain processes were genuinely subject to chance, and since language is my focus here, I cannot dismiss such talk as misguided or irrelevant.

On the other hand, the present framework is not probabilistic, thus here it is not the probability distribution over the values of a variable that is determined by the values of its parents. Rather, what is determined by its parents is, for each of its values, whether the event of its having that value is a necessity or a possibility in the Premise Semantic sense. Since I am representing variables as partitions over the set of possible worlds, “the event of a (given) variable having a (given) value” is a proposition, i.e., a cell in the partition induced by the variable in question. Whether this proposition is a necessity or a possibility given the values of the parents is determined by the premise background corresponding to the ordering source.
Some ordering sources respect the independence assumptions encoded in the network structure, others do not. Below, I formulate a condition that an ordering source must meet in relation to a given causal structure, but I do not stipulate that all ordering sources used in the interpretation of counterfactuals meet this condition. In other words, non-causal readings are not ruled out. So once again, as with the modal base, whether an interpretation is causal or not is contextually determined.

3.1 Causal independencies

The main function of the causal structure in counterfactual reasoning is to capture asymmetries under causal intervention. Consider the simple case of two variables $X, Y$, where $X$ causally precedes or dominates $Y$ (henceforth written ‘$X < Y$’). Usually in the theory of causal networks, this is depicted as in Figure 2(a). The idea is that manipulations of $X$ affect $Y$, but not vice versa.

To see intuitively what this is meant to capture, let $X$ and $Y$ stand for ‘whether it is sunny’ and ‘whether the streets are dry’, with $X < Y$ as in the figure, and suppose that in fact it is sunny and the streets are dry. An intervention to make the streets wet (say, with a water hose) would not bring about rain, whereas an intervention to make it rain (say, by seeding clouds) would result in the streets begin wet.

In the literature on Causal Networks, intervention is associated with counterfactual reasoning, which makes it a natural candidate for the interpretation of counterfactual conditionals. The prediction is that (14a) is false and (14b) is true in the situation described.

In Causal Bayesian Networks, this is captured by the ‘do’-operator with an associated special update rule, intended to model external interventions on causal processes (Pearl, 2000). Simply put, $X$ is “cut loose” from its causal parents before its value is updated. This ensures that the update has no repercussions for $X$’s parents or other non-descendants. (Technically, this is a consequence of the way in which standard belief propagation algorithms are guided by the network structure.)

These intuitions crucially depend on the assumption that the dependency between the variables is causal. The wetness of the street does not make it rain, although of course it can serve as evidence of rain.

The word ‘still’ in (14) helps bring out the intended reading, but it is optional and its semantic contribution is orthogonal to the truth conditions I aim to capture.
(14)  
   a. If it were raining, the streets would (still) be dry.  \[false\]  
   b. If the streets were wet, it would (still) be sunny.  \[true\]  

I take it to be uncontroversial that the sentences really do have readings under which the judgments are as indicated on the right-hand side in (14). This is not the only possible interpretation, especially for cases like (14b), to which I return in Section 4.1 below. For now, the goal is to capture the interpretation which leads to the judgments indicated in (14).

Recall that the variables correspond to partitions in possible-worlds semantics. The two relevant partitions in our two-variable scenario are shown in Figure 2(b). The four cells in these partitions constitute the propositions which come into play in causal reasoning in this model. Now, at each world of evaluation, exactly two of them are true. For instance, \(x \in X\) and \(y \in Y\) are both true at a world which lies in the upper left quadrant of the area in the figure, such as \(w\) in Figure 2(c). Schematically, the asymmetry in question concerns the counterfactuals in (15): At \(w\), (15a) is false whereas (15b) is true.

\[
(15)  
   a. If were \(\overline{x}\), would (still) be \(y\).  \[false \text{ at } w \in xy\]  
   b. If were \(\overline{y}\), would (still) be \(x\).  \[true \text{ at } w \in xy\]  

Intuitively, contrasts of this kind are about truths that speakers “hold on to” in exploring counterfactual scenarios: The truth of \(x\) at \(w\) is shielded from the supposition that \(\overline{y}\), but the truth of \(y\) is affected by the supposition that \(\overline{x}\). Such asymmetries are encoded in the causal structure, represented here by the premise background \(f\).

To cast these ideas in a more precise form, I start by defining the notion of a “causal structure” relative to a non-empty set \(W\) of possible worlds.

**Definition 14** (Causal structure)  
A causal structure for \(W\) is a pair \(C = (U, <)\), where \(U\) is a set of finite\(^{23}\) partitions on \(W\) and \(<\) is a directed acyclic graph over \(U\).

I assume that all partitions in \(U\) are bipartitions. Nothing hinges on this assumption, but it simplifies the exposition. I write \(X, Y, \ldots\) to refer to these bipartitions, and \(x, \overline{x}, y, \overline{y}, \ldots\) for their respective cells. I use boldfaced variable names like ‘\(X\)’ to refer to finite sets of partitions, and ‘\(x\)’ to range over settings of \(X\) – formally, \(x\) is a set of propositions containing exactly one cell from each partition in \(X\).\(^{24}\)

\(^{23}\)I ignore the case of continuous variables.  
\(^{24}\)Notice that \(x\) may contain fewer members than there are partitions in \(X\), since distinct partitions in \(X\) may agree on one or more cells. Such multiply introduced cells are collapsed in \(x\).
In this paper I discuss the evaluation of counterfactuals relative to total settings of all variables in the given causal structure. In other words, there are no unobserved causally relevant variables. This is a special case; in reality, conditionals of all stripes are routinely used and interpreted under uncertainty about relevant causal facts. But such uncertainty introduces an epistemic dimension which takes us beyond the scope of the present investigation.

Given a partition $X$ and world $w$, I refer to the cell in $X$ containing $w$ (i.e., the proposition true at $w$) as ‘$[w]_X$’. I next define two sets of propositions: The causally relevant propositions in the model are the various cells in the partitions in the causal structure. Among these, the causally relevant truths at a given world $w$ are those that are true at $w$. For the above toy example, these two sets are shown in Figure 2(b) and 2(c), respectively.

**Definition 15** (Causally relevant propositions)
For any set $U$ of partitions on $W$ and world $w \in W$:

a. The set $\Pi^U$ of causally relevant propositions is the set of all cells of all partitions in $U$.

b. The set $\Pi^U_w$ of causally relevant truths at $w$ is the set of causally relevant propositions that are true at $w$.

Henceforth I omit the superscript on $\Pi^U$ because it will be clear from the context which causal model is being talked about.

For the purposes of interpreting counterfactuals, I postulate a causal premise background $f_c$ whose value at each world of evaluation is built up from the causally relevant truths at that world. Specifically, I stipulate that $f_c(w)$ structures $\Pi_w$ in the sense discussed above (i.e., it consists of subsets of $\Pi_w$ including $\Pi_w$ itself). This means that $f_c$ is realistic in that no premise set in $f_c(w)$ contains any proposition that is false at $w$. There is also an intuitive sense in which $f_c(w)$ encodes a notion of similarity to $w$: Since $w \in \bigcap X$ for all $X \in f_c(w)$, increasing sequences of premise sets have the effect of “zooming in” on $w$. However, it is not required that $f_c(w)$ uniquely identify $w$, since the maximal member of $f_c(w)$ is $\Pi_w$ and it is not required that $\bigcap \Pi_w$ contain just $w$.

Aside from $\Pi_w$ itself, which subsets of $\Pi_w$ should be in $f_c(w)$? The first impulse might be to include all of its subsets, that is, in effect, all ways of adding (zero or more) propositions from $\Pi_w$ to the antecedent while maintaining consistency. It is easy to see, however, that this simple rule would fail to account for asymmetries of the kind in (15). Suppose as before that $X$ and $Y$ are both true at $w$, so $\Pi_w = \{x, y\}$, and thus $f_{c_p}(w)$ is as in (16a). The resulting premise sequences with the two antecedents $\overline{Y}$ and $\overline{X}$ are given in (16b,c).
(16)  
   a.  $f_p(w) = \{\emptyset, \{x\}, \{y\}, \{x, y\}\}$  
   b.  $\text{Prem}(f_p[\bar{y}](w)) = \{\bar{y}, \bar{y}.x\}$  
   c.  $\text{Prem}(f_p[\bar{x}](w)) = \{\bar{x}, \bar{x}.y\}$  

   With antecedent $\bar{y}$ we obtain the single maximal premise set $\bar{x}.x$, so (15b) comes out true, as desired. However, by the same token, with $\bar{x}$ the maximal premise set is $\bar{x}.y$, therefore (15a) is also predicted to be true, which is wrong.

   The problem is that the causal structure is not taken into account. To make the interpretation sensitive to the asymmetric relationship between $X$ and $Y$, I impose a restriction on the subsets of $\Pi_w$ that may enter the premise sets.

   First, two bits of terminology. (i) A variable $X$ is set in a set of propositions $P$ iff exactly one of $X$’s cells is in $P$. (ii) $X$ is a descendant of $Y$ iff there is a path from $Y$ to $X$ of zero or more steps along the direction of causal influence.

   **Definition 16** (Closure under ancestors)

   Let $P$ be a consistent set of propositions and $\langle U, \prec \rangle$ a causal structure. A subset $P'$ of $P$ is **closed under ancestors in** $P$ iff for all $X, Y \in U$ such that $X$ is a descendant of $Y$ and both are set in $P$, if $X$ is set in $P'$, then $Y$ is also set in $P'$.

   Closure under ancestors constrains the available premise sets: Whenever a proposition is barred from membership on pain of inconsistency, its causal descendants are also barred.

   (17)  
   $f_c(w) := \{X \subseteq \Pi_w \mid X \text{ is closed under ancestors in } \Pi_w\}$

   Consider what this definition comes down to in the two-variable example above. The four causally relevant propositions induced by the partitions $X$ and $Y$ are listed in (18a). Relative to a world $w$ at which both $x$ and $y$ are true, the set of causally relevant truths is as in (18b), and the results of revising this set with different propositions are listed in (18d) and (18e).

   (18)  
   a.  $\Pi = \{x, \bar{x}, x, \bar{x}, y, \bar{y}\}$  
   b.  $\Pi_w = \{x, y\}$  
   c.  $f_c(w) = \{\emptyset, \{x\}, \{x, y\}\}$  
   d.  $\text{Prem}(f_c[\bar{x}](w)) = \{\bar{x}\}$  
   e.  $\text{Prem}(f_c[\bar{y}](w)) = \{\bar{y}, \bar{y}.x\}$

   The crucial point is that $\{y\}$ is not closed under ancestors (therefore absent in (18c)), whereas $\{x\}$ is so closed (there present in (18c)). Consequently, $\bar{x}.y$, though logically consistent, is not an admissible premise sequence, whereas $\bar{y}.x$ is.

   25Formally, $X$ is a descendant of $Y$ iff $Y \prec^* X$, where $\prec^*$ is the reflexive transitive closure of $\prec$. Note that this means that every variable is a descendant of itself. In this I follow Spirtes et al. (2000).
Consider again the counterfactuals in (15), repeated here as (19). It is easy to see that the premise sequences in (18d) and (18e) correctly predict the judgments.

(19) a. If were $\bar{x}$, would (still) be $y$. [false]
b. If were $\bar{y}$, would (still) be $x$. [true]

This is progress. It is not yet the full story, however. Without an appeal to the dependencies between the variables, the account does not make plausible predictions even in simple cases. I turn to this in the next subsection.

### 3.2 Causal dependencies

Consider the following example from Goodman (1947).

(20) If that match had been scratched, it would have lit.

Whether the match lights depends not only on the scratching, but also on other factors, such as the presence or absence of oxygen. Let us focus on this one additional factor. Figure 3 shows the dependencies.\(^{26}\) Suppose the crucial parameters are as listed in (21). Thus at $w$ the match was not struck, oxygen was present, and the match did not light. The premise sequences derived are as shown in (21d). This ensures that (22) comes out true, as it should.

(21) a. $\Pi = \{s, \bar{s}, o, \bar{o}, l, \bar{l}\}$
b. $\Pi_w = \{\bar{s}, o, \bar{l}\}$
c. $f_c(w) = \{\emptyset, \{\bar{s}\}, \{o\}, \{s, o, \bar{l}\}\}$
d. $\text{Prem}(f, [s](w)) = \{s, s.o\}$

(22) If that match had been scratched, there would (still) have been oxygen.

However, the truth of (20) is not yet accounted for. The exclusion of $\bar{l}$ from the premise sequences is not sufficient to ensure that $l$ comes out as a necessity. This is correct inasmuch as the truth of $l$ is not “carried over” from the world of evaluation. Rather, $l$ is inferred “afresh” from the causal premise sequences (with

\(^{26}\)The variable names are straightforward: $S = \text{‘whether the match is struck’}$; $O = \text{‘whether there is oxygen’}$; $L = \text{‘whether it lights’}$.
the antecedent) in combination with assumptions about what is the most likely or normal course of events.

This kind of inference is encoded in the second premise background, $g_\ell$, which encodes the dependencies between the variables.\(^{27}\) It corresponds to what in the Kratzerian tradition is called a stereotypical ordering source, one which represents what is likely or normally the case. I do not require $g_\ell$ to be realistic: It is possible that the world of evaluation happens to be an unlikely one by its own criteria. The idea is that much like a prior probability distribution, $g_\ell$ holds general information as to what is more likely to be the case than not. Mutually inconsistent premise sets in $g_\ell(w)$ describe incompatible states of affairs that are incomparable in terms of likelihood, but may still yield unequivocal predictions in combination with premise sets from the modal base $f$.

I also do not require the premise sets in $g_\ell(w)$ to be composed solely of causally relevant truths at $w$, or for that matter of causally relevant propositions in the model. However, I do impose a constraint on the relationship between the causal structure and $g_\ell$. The (causal) Markov Condition stipulates that for any variable $X$ in the structure, given $X$’s immediate causal ancestors, $X$ is independent of its (other) non-descendants. This condition is imposed as a constraint on admissible ordering sources relative to a given causal structure. In order to get an intuitive feel for what it amounts to, it is useful to pause for a moment to reflect on the idea behind (conditional) independence in general.

Consider first two arbitrary partitions $X,Y$ and a consistent set of propositions $M$ that is disjoint from both. The idea is that $X$ is conditionally independent of $Y$ given $M$ under $g(w)$ iff for each pair of cells $x \in X$ and $y \in Y$ such that $y$ is consistent with $M$, the following holds: $x$ is a necessity (possibility) relative to Prem$\{M \ast g(w)\}$ iff $x$ is a necessity (possibility) relative to Prem$\{M \cup \{y\} \ast g(w)\}$. Intuitively, the idea is that adding information about $Y$ to $M$ (thus obtaining $M \cup \{y\}$ for some $y \in Y$) does not affect the status of any of $X$’s values.\(^{28}\)

This basic notion is generalized in two directions. The first is to independence of $X$ from a set $Y$ of variables, rather than just a single variable $Y$. In this case the status of $X$’s values must remain unaffected by adding information about any subset of $Y$. Let a partial setting of $Y$ be subset (not necessarily proper) of a setting of $Y$ – i.e., a set of propositions containing at most one cell

\(^{27}\)In much of the literature on causal independence and counterfactuals, this is done by adding a probabilistic component (Hiddleston, 2005; Kaufmann, 2005a, 2009, and others).

\(^{28}\)There is an obvious connection to probabilistic independence. One way to define the latter is to say that for all values $x_i$ of $X$ and $y_j$ of $Y$, the unconditional probability $P(x_i)$ equals the conditional probability $P(x_i|y_j)$. In this case it is the probability of each value of $X$ that is preserved under refinements.
Figure 4: A violation of conditional independence of $X$ from $Y$ given $Z$. Rectangles represent the cells of the partitions induced by the consistent settings $z_i$ (left) and $z_i \cup y_j$ (right). The shading of each cell $\varepsilon$ indicates whether $x$ is a necessity (black), only a possibility (gray), or neither (white) relative to $\text{Prem}((\{\varepsilon\} \cup \varepsilon, g(w))$. 

from each member of $Y$. Suppose $M$ is disjoint from both $X$ and all partitions in $Y$. Now $X$ is conditionally independent of $Y$ given $M$ under $g(w)$ iff for all cells $x \in X$ and partial settings $y$ of $Y$ such that $M \cup y$ is consistent, $x$ is a necessity (possibility) relative to $\text{Prem}((M) \cup g(w))$ iff $x$ is a necessity (possibility) relative to $\text{Prem}((M \cup y) \cup g(w))$.

The second generalization is to conditional independence given a set $Z$ of partitions, rather than a set $M$ of propositions. This is straightforwardly done by universal quantification over the consistent settings of $Z$: The independence must hold for all of them. The formal definition is as follows.

**Definition 17** (Conditional independence)
Let $g$ be a premise background, $w$ a possible world, and $U$ a set of partitions. For any $X \in U$ and disjoint sets $Y, Z \subseteq U$ not containing $X$: $X$ is conditionally independent of $Y$ given $Z$ under $g(w)$ iff for all cells $x \in X$, partial settings $y$ of $Y$, and settings $z$ of $Z$ such that $y \cup z$ is consistent, $x$ is a necessity (possibility) relative to $\text{Prem}((\{z\} \cup g(w))$ iff $x$ is a necessity (possibility) relative to $\text{Prem}((\{z \cup y\} \cup g(w))$.

Figure 4 gives a simplified illustration of a case in which $X$ is not independent of $Y$ given $Z$. Here, $Z$ has four consistent settings, labeled $z_1$ through $z_4$. The left-hand side of the figure shows the partition induced by these settings: Each cell $z_i$ comprises the set of worlds at which all propositions in the corresponding setting are true. The shading of each cell $z_i$ indicates the status of one of $X$’s values, say $x$, relative to $\text{Prem}((\{z_i\} \cup g(w))$: It is a necessity relative to $\text{Prem}((\{z_1\} \cup g(w))$ and $\text{Prem}((\{z_3\} \cup g(w))$; a possibility (but not a necessity) relative to $\text{Prem}((\{z_4\} \cup g(w))$; and neither a possibility nor a necessity relative to $\text{Prem}((\{z_2\} \cup g(w))$. In addition, let there be two settings of $Y$, labeled $y_1$ and $y_2$,
such that all eight pairs of settings $z_i, y_j$ are mutually consistent. This gives rise to the eight cells on the right-hand side of the figure, each corresponding to the result of adding one setting of $Y$ to a setting of $Z$. The definition of independence requires that the status of $x$ be preserved under all such refinements. But notice that $x$ is a necessity relative to $\text{Prem}((z_1 \ast g(w))$ and not relative to $\text{Prem}((z_1 \cup y_2) \ast g(w))$. This is sufficient to conclude that $X$ is not independent of $Y$ given $Z$ under $g(w)$.

Figure 4 is a simplification in two respects. First, the status of only one cell of $X$ is indicated; in fact, the status of all cells of $X$ must be preserved under refinements in the manner indicated. Second, only a single refinement by the bipartition $\{y_1, y_2\}$ is given. In general, different subsets of $Y$ may yield different refinements. The general definition requires that all ways of carving up the cells of $Z$ using variables in $Y$ preserve the status of $x$ under $g(w)$.

In stating the Markov Condition, I use the following notation. For any partition $X$ in the causal structure, I write ‘$pa(X)$’ for the set of $X$’s parents, and ‘$de(X)$’ for the set of $X$’s descendants. With these notions in place, I define the causal Markov Condition as conditional independence from all non-descendants, given the parents:

**Definition 18 (Causal Markov Condition)**

Let $C = \langle U, \prec \rangle$ be a causal structure and $g$ a premise background. $g$ satisfies the Markov Condition relative to $C$ if and only if for all $w \in W$ and $X \in U$, $X$ is conditionally independent of $U \setminus (de(X) \cup pa(X))$ given $pa(X)$ under $g(w)$.

In the following, I assume throughout that the ordering source obeys the causal Markov condition relative to the causal model in question.

Returning to the example of Goodman’s match, I assume that the ordering source supplies the information that if the match was struck in the presence of oxygen, then it (most likely) lit, whereas it (most likely) did not light otherwise. This can be encoded in various ways. What is required is a set of propositions which makes the lighting of the match a necessity when combined with a premise set that contains or entails both $s$ and $o$, and an impossibility when combined with a premise set that contains or entails either \( \overline{s} \) or \( \overline{o} \) (or both). One concrete but also simplistic way to ensure this is to use just one proposition, namely the set of worlds at which the material biconditional $(s \land o) \leftrightarrow l$ is true.

More generally, one way to represent the dependencies between a variable $X$ and its parents $pa(X)$ in compliance with the Markov Condition is as follows. For each setting $y$ of $pa(X)$ and cell $x \in X$, to encode the fact that $x$ is likely given $y$, let the ordering source provide the proposition characterized by the material con-
In this sense, the biconditional \((s \land o) \leftrightarrow l\) is really just an abbreviation, for the sake of clarity, of the four conditionals \((s \land o) \rightarrow l\), \((\neg s \land o) \rightarrow \neg l\), \((s \land \neg o) \rightarrow l\), and \((\neg s \land \neg o) \rightarrow \neg l\). The restriction of the antecedents of these conditionals to settings of \(X\)'s parents enforces the Markov Condition. This is but one way of ensuring that \(l\) is a necessity given \(s\) and \(o\). As mentioned above, there is in general no reason to assume that the ordering-source propositions must be drawn from the causally relevant propositions, or, for that matter, that they must be the denotations of sentences in the language. However, for simplicity in this paper, I stick to this way of encoding defeasible conditional preferences.

Suppose, then, that the proposition \((s \land o) \leftrightarrow l\) is provided by the ordering source in our toy example. To flesh things out a bit, suppose further that the ordering source also provides the proposition \(o\), that oxygen was present. Then the values of the parameters for the interpretation of (20), repeated here as (24), are as given in (23). The premise sets are listed in the Hasse diagram in Figure 5(a).

Figure 5: Premise sequences for (24) and (26) at worlds \(w\) and \(v\).

29The notation is sloppy. Recall that I use sentences \(\varphi\) and their denotations \(V(\varphi)\) interchangeably. In set-theoretic terms, the proposition is \(\text{\overline{\bigcap y \cap x}}\).
(24) If that match had been scratched, it would have lit.

The example in (23) also illustrates the assumption that the likelihood background may be opinionated on causally relevant propositions, as is the case in (23) with respect to \( o \). In such cases, what \( g_r(w) \) prescribes may or may not be in accordance with the facts at \( w \). If there is a conflict, then the facts (preserved in \( f_c(w) \)) override the biases in \( g_r(w) \). To illustrate, consider the evaluation of (24) relative to the same likelihood assumptions, but at a world \( v \) that is just like \( w \) except that at \( v \) there was no oxygen. The causal background at \( v \) is shown in (25b). The likelihood background contributes the same propositions at \( v \) as at \( w \) (25d). This time the bias for the presence of oxygen is trumped by the fact that there was none at \( v \). The premise sequences are shown in Figure 5(b). In (25e) I only show the maximal premise sequence to avoid clutter. As a result, the counterfactual in (24) is false; in fact, at \( v \) we obtain the stronger prediction that (26) is true.

(25) a. \( \Pi_v = \{ \bar{s}, \bar{o}, \bar{l} \} \)
    b. \( f_c(v) = \{ \emptyset, \{ \bar{s} \}, \{ \bar{o} \}, \{ \bar{s}, \bar{o}, \bar{l} \} \} \)
    c. \( \text{Prem}(f_c[s](v)) = \{ s, s. \} \)
    d. \( g_r(v) = \{ \emptyset, \{ o \}, \{ (s \land o) \leftrightarrow l \}, \{ o, (s \land o) \leftrightarrow l \} \} \)
    e. \( \text{max Prem}(f_c[s] \ast g_r)(v)) = \{ s, \bar{s}, \bar{s} \} \}

(26) If the match had been scratched, it would not have lit.

These considerations highlight a more general property of the framework, briefly alluded to above: The premise sets contributed by the causal background work towards similarity with the world of evaluation (thus \( \bar{s} \) is preserved in (25e)). The likelihood background, in contrast, may introduce dissimilarity if the world of evaluation is unlikely by its own criteria (as in (25d)). If such a conflict arises, the facts win.

4 Backtracking and explanation

The ingredients are in place now, but the account is still rather limited. In this section I extend the basic analysis to some more complicated types of examples.

4.1 Backtracking inferences

The interpretation of counterfactuals so far introduced proceeds, at least for those with false antecedents, by intervention in a maximally local sense: All non-descendants of the conditional antecedent, including its parents, are left un-
touched.\textsuperscript{30}

But are counterfactuals always interpreted in this way? There are good reasons to believe that this association is more a soft preference than a hard rule. A number of experimental studies cast doubt on the assumption of strictly local intervention (Sloman and Lagnado, 2004; Rips, 2009; Dehghani, Iliev, and Kaufmann, 2012). The reader is referred to those studies for details; here I only illustrate the issue with a well-known example from the philosophical literature. Lewis (1979) discussed the following scenario, originally due to Downing (1959).

Jim and Jack quarreled yesterday, and Jack is still hopping mad. We conclude that

\begin{align*}
(27) & \quad \text{If Jim asked Jack for help today, Jack would not help him.} \\
\end{align*}

But wait: Jim is a proud fellow. He never would ask for help after such a quarrel; \textit{if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday}. In that case Jack would be his usual generous self.

\begin{align*}
(28) & \quad \text{So if Jim asked Jack for help today, Jack would help him after all.} \\
\end{align*}

The main point here concerns the difference between (27) and (28). Readers generally agree that the story is plausible and both counterfactuals are true in their respective contexts. However, it is also intuitively clear that the two would contradict each other when taken out of context. What the example shows, then, is that counterfactuals are context-dependent in a serious way, so much so that the few intervening statements can induce a radical shift which reverses readers’ judgments. But what exactly is the contextual factor at work here, and where in the account should it be represented?

Lewis (1979) used the example to delineate the scope of his counterfactual analysis of causality. The context for the judgment in (28) is created by the counterfactual in the intervening paragraph that is emphasized (by me) in the scenario. The emphasized sentence is a \textbf{backtracking} counterfactual, instantiating reasoning from effect to cause, for which the clean correspondence between counterfactual and causal asymmetries that Lewis investigated in his paper breaks down. My present concerns are different from Lewis’s, of course. I take a causal model for

\textsuperscript{30}If the antecedent is true at the world of evaluation, the descendants are also untouched. This assumption is often made in the philosophical literature (Stalnaker, 1968; Lewis, 1973, and many more) as well as in Kratzer’s (1981a; 1989) analysis. I adopt it here without further discussion.
The kind of reading under which (28) comes out true has been called a backtracking resolution (for an otherwise non-backtracking counterfactual) by Arregui (2005a). I adopt this term for my purposes. It is important to keep in mind that the presence of the emphasized backtracker in the prose between (27) and (28) is not essential for the availability of the backtracking resolution: The resolution itself is always contextually available and only made salient by the explicit mention of the backtracker. My concern in this section is with the inference involved, rather than the linguistic analysis of the explicit backtracker.

To get things started, let us fix a few parameters. There are three variables, depicted in Figure 6, indicating whether there was quarrel (Q), whether Jim asks Jack for help (A), and whether Jack helps Jim (H). I take the causal relations to be as indicated by the arrows. The crucial parameters are listed in (29). (The assumption that \( \overline{h} \) is true at the world of evaluation is added for completeness. It does not affect the interpretation.)

\[ \begin{align*}
\text{(29) } & \text{a. } \Pi_w = \{q, \overline{a}, \overline{h}\} \\
& \text{b. } f_c(w) = \{\emptyset, \{q\}, \{q, \overline{a}\}, \{q, \overline{a}, \overline{h}\}\} \\
& \text{c. } \text{Prem}(f_c[a](w)) = \{a, \overline{a}, q\} \\
& \text{d. } g_c(w) = \{\emptyset, \{q \rightarrow \overline{a}\}, \{(\overline{q} \land a) \leftrightarrow h\}, \{q \rightarrow \overline{a}, (\overline{q} \land a) \leftrightarrow h\}\} \\
& \text{e. } \text{max Prem}(f_c[a] \ast g_c)(w) = \{a, \overline{a}, (\overline{q} \land a) \leftrightarrow h\}\end{align*} \]

(27) If Jim asked Jack for help today, Jack would not help him.

The premise sequences obtained from the hypothetically updated modal base alone are shown in (29c). The premise sequences for the counterfactual are obtained by combining elements of (29c) with elements of (29d). The set of all premise sequences is shown in the Hasse diagram in Figure 7(a); (29e) only lists the

\[ \begin{align*}
\text{max Prem}(f_c[a] \ast g_c)(w) = \{a, \overline{a}, (\overline{q} \land a) \leftrightarrow h\}\end{align*} \]

Figure 6: Causal structure for (27) and (28)

\[ \begin{align*}
\text{granted and ask whether and how it can help us in the analysis of counterfactuals in general, not just “forward” ones in the direction of causality.}
\end{align*} \]
single maximal premise sequence. It entails \( \overline{h} \), thus the truth of (27) is accounted for.

But (28) is true as well, under a different interpretation. The basic intuition that I would like to pursue is that the counterfactual supposition that Jim would ask Jack is surprising given the fact that there was a quarrel yesterday, and that speakers endorse (28) upon finding in the absence of the quarrel a plausible (and likewise counterfactual) explanation for the supposed asking.\(^{32}\) The place where I propose to model the difference is in the selection of premise sequences from the sequence structure. So far I have imposed the sole requirement from Definition 9 that they be consistent, where consistency was understood as logical consistency. I take examples like the Jim/Jack scenario to show that a stronger notion of consistency can play a role as well, viz. one defined in terms of possibility (in the Premise Semantic sense) relative to an ordering source.

Formally, I define a notion of likelihood premise structure as follows: Among all the premise sequences \( p. F. G \) for the conditional (i.e., all logically consistent sequences), only those are used for which the antecedent \( p \) is a possibility (in the Kratzerian sense) given \( F \).

**Definition 19** (Likelihood premise structure)

Let \( f, g \) be premise backgrounds, \( w \) a possible world, and \( p \) a proposition. The **likelihood premise structure** \( \text{Prem}^\ell((f[p] \ast g)(w)) \) is the set of sequences \( p. F. G \in \text{Prem}((f[p] \ast g)(w)) \) with \( F \in f(w), G \in g(w) \), such that \( p \) is a possibility relative to \( \text{Prem}((f) \ast g)(w)) \).

In Lewis’s example discussed above, this interpretation leads to a different outcome. Premise sequences whose modal-base component entails \( q \) are no longer admissible, since it is impossible (given the ordering source) for Jim to ask for help given that the quarrel occurred. Most of the parameters are as in (29) above, except for the premise structure, whose sole maximal element is shown in (30). It entails both that there was no quarrel and that Jack helps Jim. Figure 7(b) shows the likelihood premise structure in a Hasse diagram.

\[(30) \quad \text{max} \text{Prem}^\ell((f, [a] \ast g, \ell)(w)) = \{a..(q \rightarrow \overline{a})(\overline{a} \wedge a) \leftrightarrow h)\}
\]

How do speakers and hearers determine which notion of consistency is used on any particular occasion? I have no answer to this question at this point; however, examples like the one discussed in this section make a strong argument that the distinction captures a true variation along what I believe to be a contextual variation.

\(^{32}\)See Dehghani et al. (2012) for an outline of this idea for counterfactuals in the framework of Gärdenfors’s (1988) theory of explanation, and Bennett (2003); Kaufmann (2004) for a related approach to indicatives.
4.2 Disjunctive premise sets

All of the examples discussed so far involved interventions on causal variables with at most one parent. The case of multiple parents is worth considering because it highlights a feature of the current account viz. its treatment of what might be called “disjunctive” premise backgrounds.\footnote{A more appropriate label might be inquisitive, alluding to a particular strand of research in formal semantics and pragmatics (see the papers at \url{https://sites.google.com/site/inquisitivesemantics/Home}). I use the more widely comprehensible term “disjunctive” here for ease of exposition and reserve a more extensive comparison with the particular framework of Inquisitive Semantics for future work.}

Consider again the causal structure behind Goodman’s match example, this time with regard to the examples in (31).

(31)  a. If the match had not lit, it would (still) have been struck.
     b. If the match had not lit, there would (still) have been oxygen.

Judgments on these sentences are variable and context-dependent; the goal is to capture at least some of that context-dependence. Suppose the only ordering-source proposition concerns the dependency between the variables (the match lights if and only if it is struck and oxygen is present) and no other assumptions are made. In this case, both (31a) and (31b) should come out false. If, on the other
hand, the ordering source also encodes (in addition to the aforementioned dependency between the variables) the assumption that oxygen is stereotypically present, then (31a) should come out false whereas (31b) is true.

The value of the modal base at \( w \) is as before, repeated here as (32a). The update with the antecedent \( l \) results in (32b).

\[
\begin{align*}
\text{(32)} & \quad \text{a. } f_c(w) = \emptyset, \{s\}, \{o\}, \{s, o, l\} \\
& \quad \text{b. } f_c(l)(w) = \{l, l.s, l.o\} \\
& \quad \text{c. } g_1(w) = \emptyset, \{(s \land o) \leftrightarrow l\} \\
& \quad \text{d. } g_2(w) = \emptyset, \{o\}, \{(s \land o) \leftrightarrow l\}, \{o, (s \land o) \leftrightarrow l\}
\end{align*}
\]

The premise sequences obtained with these parameters are shown in Figure 8. (Sequences of the form \( l.s.o.G \) are omitted because they cannot be consistent regardless of \( G \), hence to not affect the interpretation of modal expressions.) The graph on the left shows the set of premise sequences obtained under the ordering source \( g_1 \) at \( w \), listed in (32c). It turns out that in this case the premise structure and the likelihood premise structure are identical. There are two maximal premise sequences, listed in (33).

\[
\text{(33)} \quad \text{max Prem}(f_c[l] * g_1(w)) = \{l.s.(s \land o) \leftrightarrow l), l.o.(s \land o) \leftrightarrow l\}
\]

The first sequence in (33) entails that the match was struck in the absence of oxygen, whereas the second one entails that oxygen was present and the match was not struck. Since both are maximal, neither of the counterfactuals in (31) comes out true, which is as desired.

Things are different with the second ordering source \( g_2 \), which at \( w \) includes
the information that oxygen is (stereotypically) present.34 In this case, there is a difference between the ordinary premise sequences and the likelihood premise sequences. Since the ordering source makes it unlikely that the match would have lit given that it was struck, the sequences of the form \( \overline{I}.s.G \) are not eligible for membership in the likelihood premise structure. As the figure shows, this time the likelihood premise sequences are in fact upper-bounded with the single maximal element in (34), which entails that oxygen was present and the match was not struck. Hence the counterfactual in (31b) is true whereas the one in (31a) is false, again as desired.

(34) \text{max} \text{Prem}((f[\overline{I}] \ast g_2)(w)) = \{ \overline{I}.o.((s \land o) \leftrightarrow l) \}

As a final note, it is worth pointing out that disjunctive modal bases may also arise from disjunctive antecedents. For this to work, it has to be assumed that disjunctions are represented as sets of propositions. This is in fact the case in the framework of Inquisitive Semantics. This is not the place for an adequate description of this framework (see Groenendijk, 2009; Groenendijk and Roelofsen, 2009, 2010, among others). Suppose, merely, that an antecedent of the form 'if(\( r \) or \( s \))' is represented as \([\{r\}, \{s\}]\), and consider for concreteness the premise background \( f_c \) from (11) above. Let us assume that all four propositions involved are mutually consistent.

(11) \( f_c(w) = \{\emptyset, \{p\}, \{p, q\}\} \)

The set of premise sequences obtained by pairing this set with \([\{r\}, \{s\}]\) is shown in Figure 9. What is important about these sequences is that there are two maximal

---

34The graph in Figure 8(b) could be simplified by collapsing semantically equivalent premise sequences. I refrain from doing so here because the redundancy does not affect the interpretation of modal claims and the structure of the premise sequences is more transparent if the redundant ones are listed.
ones, $r.pq$ and $s.pq$. This means that for a proposition $t$ to be a necessity relative to this sequence set, $t$ must be a necessity entailed by both $r.pq$ and $s.pq$; in other words, the counterfactual in (35a) is equivalent to the one in (35b).

\[ \text{(35) a. If were 'r or s', would be t.} \]
\[ \text{b. If were 'r', would be t and if were 's', would be t.} \]

This is an encouraging result; however, a full exploration of counterfactuals with complex antecedents must wait for another occasion.

5 Conclusions

I hope to have demonstrated that a Premise Semantic account of the causal inferences that tend to enter the interpretation of counterfactuals is not only possible, but in fact fairly straightforward. Moreover, the particular variant of Premise Semantic in which the present proposal is implemented opens up promising avenues for the analysis of complex (e.g., disjunctive) antecedents. The connection to Inquisitive Semantics mentioned above suggests an avenue towards an analysis of more complex constructions, such as conditional questions (36a) a so-called un-conditionals (36b) in the same framework.

\[ \text{(36) a. If the match had not lit, would it have been struck?} \]
\[ \text{b. Whether the match had been struck or not, it would not have lit.} \]

Another open question concerns the relationship between causal relations and Kratzer’s “lumping” which we briefly mentioned in Section 2.2.2 above. Both are ways to account for the fact that facts that are “given up” in the interpretation of counterfactuals take others in their wake while leaving others undisturbed. It would be desirable to be able to point to a single notion which is uniformly at play in all counterfactuals. Causality does not seem to be general enough to play that role, for there are counterfactuals which do not clearly rely on causal relations. This concerns, for instance, the difference between what Kratzer calls “generically” and “accidentally” true quantificational statements, like the one in (37a). On the former reading it implies (37b), while on the latter reading it does not.

\[ \text{(37) a. All coins in my pocket are silver.} \]
\[ \text{b. If I had another penny in my pocket, it would be silver.} \]

In order to account for such examples by appealing to causality, one would have to postulate a causal dependence between universal statements and their verifying instances, whose direction presumably would depend on the reading involved. In

34
other words, the true of (37a) would either “cause” the truth of individual propositions such as 'that I have coin $x$ in my pocket (in addition)', or be (jointly) “caused” by them. I am unsure whether this would make for a sensible analysis without a redefinition of the intuitive notion of causality. As of now I doubt that causal relations are relevant to such examples. Perhaps, then, lumping is the more general notion and causal relations are among the factors giving rise to it. In view of the above sections, the slogan would be that effects lump their causes. This may be a worthwhile direction to explore.

Acknowledgments
[to be added]

References


