A Basic Analysis of Conditionals

Iddo Lev
Paper for Semantics Seminar Ling233A
Spring 2003

1 Introduction

The meaning of conditionals in natural language is highly context dependant, and it is difficult to find an analysis for them which will work in all contexts. The idea of this paper is to examine their meaning and usage in the particular domain of logic puzzles. The proposed analysis takes into account a context of possible worlds, but is basic in the sense that it does not resort to similarity relations between the worlds, as is common in theories of conditionals. Still, the domain is not trivial, and its analysis needs to take into account usage conditions, non-monotonic reasoning, and the inherent nature of natural language to rely implicitly on background knowledge. The analysis will then be compared to other theories of conditionals and to common case studies.

2 Conditionals in Logic Puzzles

A logic puzzle, such as those in the old analytic section of the Graduate Record Exam (GRE) and the Law School Admission Test (LSAT), consists of a preamble and a few multiple-choice questions (see Figure 1 for a typical example). The preamble describes a scenario consisting of a small number of objects that stand in certain relations to each other, and provides constraints on the possible tuples of these relations. A question is usually phrased as a conditional that stipulates additional constraints on the scenario, and asks which of the choices is true in the restricted scenario. The preamble and the questions use the modals *must* and *can*/*may* and their negations. We would like an analysis that accounts for these conditionals and modals.

2.1 Modals

Logic puzzles can be formalized using modal first-order logic (MFOL), which is FOL enriched with the two modal operators $\Box$ (necessarily) and $\Diamond$ (possibly), and the conditional operator $\rightarrow$. As will be shown below, the conditional is not the same as the material implication $\supset$ which is defined as usual: $\psi \supset \varphi \equiv \neg\psi \vee \varphi$.

Let D be the finite domain of objects described in the puzzle. We assume that each element $d \in D$ has a unique constant $\widehat{d}$ in the language. For a formula ψ and an assignment g, we use φ^g for the formula obtained from φ by replacing every variable x

Preamble: Six sculptures – C, D, E, F, G, and H – are to be exhibited in rooms 1, 2, and 3 of an art gallery. The exhibition conforms to the following conditions:

(1) Sculptures C and E may not be exhibited in the same room.
(2) Sculptures D and G must be exhibited in the same room.
(3) If sculptures E and F are exhibited in the same room, no other sculpture may be exhibited in that room.
(4) At least one sculpture must be exhibited in each room, and no more than three sculptures may be exhibited in any room.

Question 1: If sculpture D is exhibited in room 3 and sculptures E and F are exhibited in room 1, which of the following may be true?

(A) Sculpture C is exhibited in room 1.
(B) No more than 2 sculptures are exhibited in room 3.
(C) Sculptures F and H are exhibited in the same room.
(D) Three sculptures are exhibited in room 2.
(E) Sculpture G is exhibited in room 2.

Question 2: If sculptures C and G are exhibited in room 1, which of the following may NOT be a complete list of the sculpture(s) exhibited in room 2?

(A) Sculpture D (B) Sculptures E and H (C) ...

Adapted from [Web99].

Figure 1: Example of a Puzzle Text

with $\widehat{g(x)}$. Let $\mathcal{P}$ be the set of all atomic ground formulas. A possible world, i.e. a certain configuration of relations on the objects, can be identified with a subset of $\mathcal{P}$. The truth conditions of a formula without modals is defined as usual in FOL, with respect to a world w and assignment g of variables to elements of D:

- For atomic formula φ, $[\![\varphi]\!]_{w,g} = 1$ if $\varphi^g \in w$, and 0 otherwise.
- $[\![\psi \wedge \varphi]\!]_{w,g} = [\![\psi]\!]_{w,g} \cdot [\![\varphi]\!]_{w,g}$
- $[\![\forall x.\psi]\!]_{w,g} = 1$ if for all $d \in D$, $[\![\psi]\!]_{w,g[x \leftarrow d]} = 1$
  (where the assignment $g[x \leftarrow d]$ is like g except that x is assigned d)

The truth conditions of a formula with the modals $\Box$ and $\Diamond$ should not be expressed with respect to just one world, but rather with respect to some context of possibility, i.e. some set of worlds $C \subseteq H_0$, where $H_0 = 2^{\mathcal{P}}$ is the default context. We add C as a subscript to the cases above, and add:

- $[\![\Box\psi]\!]_{C,w,g} = 1$ if for all $w' \in C$, $[\![\psi]\!]_{C,w',g} = 1$
  $[\![\Diamond\psi]\!]_{C,w,g} = 1$ if for some $w' \in C$, $[\![\psi]\!]_{C,w',g} = 1$

Any of the subscripts of $[\![\cdot]\!]$ can be dropped whenever it is ignored (e.g. $[\![\Box\psi]\!]_{C,w,g} = [\![\Box\psi]\!]_{C,g}$). The modalities here are the same as in the modal logic *S5*, because the accessibility relation between worlds is trivial (all are accessible from all), and in particular the following equivalences hold: $\Box m\psi \equiv \Diamond m\psi \equiv m\psi$ for $m \in \{\Box, \Diamond\}$.

2.2 The Entailment Relation

The preamble can be formalized as a set Γ of MFOL sentences. A context $C \subseteq H_0$ is consistent with the preamble Γ iff $[\![\Box\psi]\!]_C = 1$ for all $\psi \in \Gamma$. Usually, a preamble constraint is already expressed with $\Box$ – this is evident in constraint (2) of Figure 1, but also in constraint (1) as "may not", formalized as $\neg\Diamond$, is equivalent to $\Box\neg$. We would like to have an entailment relation such that precisely the correct answer-choices are entailed by the preamble. We need to take into account two issues: the closed-world assumption of the preamble, and implicit background knowledge.

2.2.1 Closure of Constraints

Usually, the entailment relation for modal logic is defined to be: $\Gamma \vdash \varphi$ iff all contexts[1] consistent with Γ are consistent with φ. This relation is too conservative for drawing conclusions from the puzzle texts. Consider the puzzle in Figure 1. The preamble may be followed by a question: "Which of the following statements is true", and an answer choice: "Sculpture C may be exhibited in room 1", formalized $\Diamond exhibit(C, 1)$. Although $exhibit(C, 1)$ is consistent with the preamble, $\Diamond exhibit(C, 1)$ does not follow from the preamble according to $\vdash$, simply because $\vdash$ considers *all* contexts C consistent with the preamble, including singleton contexts with just one world, where $exhibit(C, 1)$ is false. However, the puzzle has an implicit assumption: the objects described in the text must obey the given constraints, *and no others*. Only in this sense, $\Diamond exhibit(C, 1)$ is supported by the preamble.

The solution is to define a non-monotonic, or "closure", relation $\Vdash$ based on the largest context consistent with the preamble. Let $Cxt(\Gamma) = \{C \subseteq H_0 : \forall\psi \in \Gamma.\ [\![\Box\psi]\!]_C = 1\}$ be the set of all contexts consistent with the preamble, and let $C^*_\Gamma = \bigcup Cxt(\Gamma)$.[2] Then $\Gamma \Vdash \varphi$ iff $C^*_\Gamma \neq \emptyset$ and $[\![\Box\varphi]\!]_{C^*_\Gamma} = 1$. Now indeed $\Diamond exhibit(C, 1)$ follows from the preamble according to $\Vdash$. This relation is non-monotonic, because even if $\Gamma \Vdash \Diamond\varphi$, adding another constraint to Γ may render φ impossible in the restricted context.

2.2.2 Implicit Knowledge

To find a unique answer to every question of a puzzle, background information is required beyond the literal meaning of the text. In Question 1 of Figure 1, for example, without the constraint that a sculpture may not be exhibited in multiple rooms, answers B, D and E are all correct. Human readers deduce this implicit constraint from their knowledge that sculptures are physical objects, rooms are locations, and physical objects can have only one location at any given time. To account for this implicit information, we relativize $\Vdash$ to a background set of assumptions Δ as follows: $\Gamma \Vdash_\Delta \varphi$ iff $\Gamma \cup \Delta \Vdash \varphi$.

[1] Or Kripke structures.

[2] $Cxt(\Gamma)$ is an upper semi lattice under $\subseteq$, and $C^*_\Gamma \in Cxt(\Gamma)$ is the maximal element.

Of course, identifying what Δ is or should be is a very hard AI problem, which is beyond the scope of this paper, but the relativization of $\Vdash$ will still be useful below.

It should be noted that implicit assumptions can come into play in another way as well: by limiting the context of possibilities C against which formulas are evaluated. Later we will talk about certain kinds of contexts (epistemic context E_a, hypothetical context H_a etc.), and about context shift during a discourse. This can be modelled either by restricting the context of evaluation or equivalently by adding more material to the background theory.

2.3 The Conditional

We observe that in addition to its truth conditions, a conditional sentence carries a presupposition that its antecedent is possible (in some context); and it may carry the implicature that the negation of its antecedent is possible. We consider a few examples.

2.3.1 Modal-free Antecedent

1. "If sculpture D is exhibited in room 1 then sculpture E *must* be exhibited in room 2", formalized as: $exhibit(D,1) \to \Box exhibit(E,2)$. This sentence can be evaluated, or has (classical) truth conditions, only in a context C where sculpture D may indeed be exhibited in room 1, i.e. this is true in some world of C. So its "usage conditions" are: $[\![\Diamond exhibit(D,1)]\!]_C = 1$. If the condition holds, then the sentence is true in the context just in case $\Box(exhibit(D,1) \supset exhibit(E,2))$ is.

2. "If sculpture D is exhibited in room 1 then sculpture E is exhibited in room 2". In the context of the preamble which describes constraints on legal configuration, the sentence has an implicit outscoping operator: "It must be the case that" or "Every legal configuration conforms to the constraint that", which can be formalized as $\Box$. So this ends up being the same as the previous case.

3. "If sculpture D is exhibited in room 1 then sculpture E *may* be exhibited in room 2", formalized as: $exhibit(D,1) \to \Diamond exhibit(E,2)$. This sentence has the same presupposition as above. If the presupposition holds, then the sentence is true in the context just in case $\Diamond(exhibit(D,1) \wedge exhibit(E,2))$ is.

In these cases, the antecedents do not involve modals. For such an antecedent φ, we can define the restriction of a context C to φ as $C|\varphi = \{w \in C : [\![\varphi]\!]_w = 1\}$. The truth conditions of a conditional are defined:

$$[\![\varphi \to \psi]\!]_{C,g} = \begin{cases} [\![\Box\psi]\!]_{(C|\varphi^g),g} & \text{if } [\![\Diamond\varphi]\!]_{C,g} = 1 \\ ? & \text{otherwise} \end{cases}$$

(The conditional is a modal operator, and like the other modal operators, its semantics does not depend on any particular world but rather on the entire context, so we drop the world subscript form the definition). The condition $[\![\Diamond\varphi]\!]_{C,g} = 1$ is equivalent to $C|\varphi^g \neq \emptyset$. The idea is that the conditional presupposes the possibility of its antecedent in the considered context. If the presupposition fails, the conditional does not have a classical truth value (just as "The king of France is bald" is neither true nor false).

A few things to note:

1. $\varphi \rightarrow \Box\psi \equiv \Box(\varphi \rightarrow \psi)$

2. if $C|\varphi \neq \emptyset$ then $\varphi \rightarrow \Diamond\psi$ has the same (classical) truth value as $\Diamond(\varphi \wedge \psi)$, but if $C|\varphi = \emptyset$ then the conditional has the value '?' while $\Diamond(\varphi \wedge \psi)$ has the value 0.

3. $\varphi \rightarrow \psi \not\equiv \varphi \supset \psi$ because the truth value of $\varphi \supset \psi$ depends on the particular world in which it is evaluated, whereas $\varphi \rightarrow \psi$ is a modal operator that takes into account all worlds in the context. We do have $\varphi \rightarrow \psi \Rightarrow \varphi \supset \psi$.

4. $\varphi \rightarrow \psi \not\equiv \Box(\varphi \supset \psi)$ because if $\psi = \Diamond\gamma$ then $\Box(\varphi \supset \psi) \equiv \Box(\neg\varphi \vee \Diamond\gamma) \equiv (\Box\neg\varphi) \vee \Diamond\gamma$, which is evaluated to 1 when $C|\varphi = \emptyset$, whereas $\varphi \rightarrow \psi$ gets '?'. Even when $C|\varphi \neq \emptyset$, $(\Box\neg\varphi) \vee \Diamond\gamma$ gets true if γ is true in some world of C, regardless of whether it is one of the φ-worlds. So in general, $\Box(\varphi \supset \psi)$ does not entail $\varphi \rightarrow \psi$. Only when φ and ψ do not involve modals, and when $C|\varphi \neq \emptyset$, do $\Box(\varphi \supset \psi)$ and $\varphi \rightarrow \psi$ get the same classical truth value.

5. $[\![\varphi \rightarrow \psi]\!]_C = 1$ iff $\varphi \Vdash_{\Gamma_C} \psi$ iff $\Gamma_C \cup \{\varphi\} \Vdash \psi$, where Γ_C is a set of formulas that defines C, i.e. C is the largest element of $Cxt(\Gamma_C)$. (Such Γ_C exists for any C because we are dealing with a finite domain and vocabulary).

6. $\Gamma \Vdash \varphi \rightarrow \psi$ iff $\Gamma \cup \{\varphi\} \Vdash \psi$ iff $\varphi \Vdash_\Gamma \psi$.

2.3.2 Modals in the Antecedent

The picture is more complicated if we allow modals in the antecedent. First, consider: "If sculpture D must be exhibited in room 1 then sculpture E must be exhibited in room 2" $\Box exhibit(D, 1) \rightarrow \Box exhibit(E, 2)$. In contrast to the cases above, this cannot be transformed to an equivalent form where there is only one $\Box$ or $\Diamond$ that has outermost scope. The sentence doesn't seem to have a presupposition, and its truth conditions seem to be equivalent to those of $\Box exhibit(D, 1) \supset \Box exhibit(E, 2)$, i.e. either $\Box exhibit(E, 2)$ is true in C, or $\Box exhibit(D, 1)$ is false in C. However, consider a question: "If sculpture E is exhibited in room 1, which of the following may be true?", formalized as: $exhibit(E, 1) \rightarrow \Diamond\gamma$, where γ is replaced by answer choices. That question could also be phrased as (2) "If sculpture E *must be* exhibited in room 1, which of the following may be true?", paraphrased as: "If the preamble is extended to include the constraint that sculpture E *must be* exhibited in room 1, then which of the following may be true?".

A more complicated case is the following. Suppose the puzzle describes the showing of a few films, and then asks: "In case films D and E are shown as far apart from each other as possible, which among the following would be true?", with a possible answer choice: "A is shown earlier than B". Note that the possibility referred to in the phrase "as far ... as possible" is relative to the constraints given in the preamble. We formalize "the time of X's showing" using $\pi(X)$, and the different between the showing times of x

and y using $|\pi(x)-\pi(y)|$, written in short as $f(x,y)$. If C is the context of the preamble (as defined in section 2.2.1) then the question asks whether an answer choice is true in the restricted context $C_m = \{w \in C : [\![f(D,E)]\!]_w = \delta\}$, where $\delta = \max_{w \in C}[\![f(D,E)]\!]_w$. The question can be formalized using the construct $let(x, \gamma, \varphi)$ (meaning: let x be γ in φ), where x is a variable, γ is a term, and φ is a formula. The construct is evaluated using: $[\![let(x,\gamma,\varphi)]\!]_{C,w,g} = [\![\varphi]\!]_{C,w,g[x \leftarrow [\![\gamma]\!]_{w,g}]}$. To mirror the natural language phrasing of the question, we formalize it as: $\psi_1 \rightarrow \chi = [let(s, f(D,E), [\neg\Diamond f(D,E) > s]) \rightarrow \chi]$ (where χ is a place holder for the answer choice). Here the antecedent of the conditional $\rightarrow$ includes a modal, so to evaluate the formula, we need to extend the definition of $C|\varphi$ to be $\{w \in C : [\![\varphi]\!]_{C,w} = 1\}$, which is like the original definition except that φ is evaluated w.r.t. both C and w.[3] Using this definition of restriction, and the same definition of $\rightarrow$, we get the desired result: let $r = [\![\psi_1 \rightarrow \chi]\!]_C$. r has value 'undefined' if $C|\psi_1 = \emptyset$, which is true iff for all $w' \in C$, $[\![\psi_1]\!]_{C,w'} \neq 1$. But of course for some $w' \in C$, D and E are at least as far apart from each other as in any other world of C. In fact, $C|\psi_1 = C_m$. So the presupposition of the antecedent is satisfied, and $r = [\![\chi]\!]_{C_m}$.

2.3.3 Implicature

We saw above that a question "If φ then ψ?" has the presupposition $[\![\Diamond\varphi]\!]_C = 1$. It also has the implicature $[\![\Diamond\neg\varphi]\!]_C = 1$ because if φ must be true according to the preamble, it is strange to stipulate it as an added condition for the question. That this is cancellable can be seen in the following cases, describing a line of reasoning:

- We know that φ, and if φ then ψ, therefore ψ.
  (The first phrase cancels the implicature of the second.)
- If φ then ψ. Indeed, φ. Therefore, ψ.
  (The second sentence cancels the implicature of the first.)

In contrast, the presupposition is not cancellable, even in a "proof by contradiction" argument. The following is infelicitous: "φ, because if $\neg\varphi$ then ψ. But we know that $\neg\psi$, therefore φ." Such an argument could be expressed using the *counterfactual*: "If $\neg\varphi$ were/had been true then ψ would be true" – this is discussed in section 3.2.

3 Extending the Results

The puzzle texts usually do not involve times (past and future) or counterfactuals. Still, in this section, we try to extend the framework to include these issues. In particular, we would like to give an account for all the cases in Figure 2.

[3] Note that C and not $C|\varphi$ is used to define the set. Perhaps in more complex cases we need to find a fixed point solution C^* to the equation $C^* = \{w \in C : [\![\varphi]\!]_{C^*,w} = 1\}$.

$\varphi = In(boxC, room1)$	Box C is in room 1
$\psi = In(boxE, room2)$	Box E is in room 2
$\varphi \rightarrow \psi$	If Box C is in room 1, Box E is in room 2
$P_1\varphi \rightarrow P_1\psi$	If Box C was in room 1 yesterday, Box E was in room 2 yesterday
$F_1\varphi \rightarrow F_1\psi$	If Box C is in room 1 tomorrow, Box E will be in room 2 tomorrow
$\varphi \rightsquigarrow \psi$	If Box C were in room 1 (today), Box E would be in room 2
$P_1\varphi \rightsquigarrow P_1\psi$	If Box C had been in room 1 yesterday, Box E would have been in room 2 yesterday
$F_1\varphi \rightsquigarrow F_1\psi$	If Box C were to be in room 1 tomorrow, Box E would have been in room 2 tomorrow

Figure 2: Examples of epistemic and hypothetical conditionals

3.1 Time

To deal with time, we extend the logical language to include the operators F (in the future), F_t (in the future, t time points from now), P (in the past), P_t, and $@_t$ (at time t). Let T be the set of integers, each representing a time point, which we will call *day* for simplicity (so it will be easy to talk about 'yesterday' and 'tomorrow'). Now the default context is $H_0 = T \rightarrow 2^{\mathcal{P}}$ where each world maps each time point to the set of propositions true at that point (in that world). All the definitions of the semantics $[\![\cdot]\!]$ are extended to include a subscript for a time point. It has an effect in the following cases:

- $[\![F_{t'}\psi]\!]_{C,w(t),g} = 1$ if $[\![\psi]\!]_{C,w(t+t'),g} = 1$
  Similarly with P_t.
- $[\![F\psi]\!]_{C,w(t),g} = 1$ if for some $t' \in T$ such that $t < t'$, it holds that $[\![\psi]\!]_{C,w(t'),g} = 1$
  Similarly with P.
- $[\![@_{t'}\psi]\!]_{C,w(t),g} = 1$ if $[\![\psi]\!]_{C,w(t'),g} = 1$
  $@_t\psi$ may be abbreviated as ψ^t.

Past tenses are analyzed using the P operators, and future tenses using the F operators (see Figure 2 for some examples).

3.2 Epistemic vs. Hypothetical Conditionals

The conditional $\rightarrow$ we have seen above is an epistemic conditional. A speaker can used it in cases where he or she has lack of knowledge about a given situation. The conditional presupposes that it is not impossible for the actual situation to be consistent with the antecedent. Another kind of conditional is the hypothetical, or counterfactual, conditional, which will be marked with $\rightsquigarrow$. It presupposes (or implicates?) that its antecedent is in fact false in the real world. However, in a larger context of possibilities that the speaker can imagine, which describe how the real world could have been (but isn't), the antecedent is still possible.[4] The difference can be seen in the following example:

1. (a) I don't know whether Box B is in room 1.
   (b) I know that Box B is not in room 1.

2. If Box B is in room 1, then Box C [is / must be / could be] in room 2.

3. If Box B were in room 1 (today), then Box C would be in room 2.

The idea in this example is that (2) can follow (1.a) but not (1.b). To express the idea behind (1.b)+(2), one needs to use (1.b)+(3).[5]

We suggest that the two conditionals have the same basic structure:

$$[\![\varphi \mapsto \psi]\!]_{C,t} = \begin{cases} [\![\Box\psi]\!]_{(C'|\varphi^t),t} & \text{if certain usage conditions hold} \\ ? & \text{otherwise} \end{cases}$$

where C is some context relevant to the evaluation of $\mapsto$

The two conditionals differ in their usage conditions (and in their implicatures), and in the context against which they are evaluated:

- Epistemic conditional: $\rightarrow$
  usage conditions: $[\![\Diamond\varphi]\!]_{E_a,t} = 1$
  context: E_a, all the worlds the speaker a considers as possible but has lack of knowledge regarding which of them is the actual world.

- Hypothetical conditional: $\rightsquigarrow$
  usage conditions: $[\![\Diamond\varphi]\!]_{E_a,t} = 0$ and $[\![\Diamond\varphi]\!]_{H_a,t} = 1$
  context: H_a, all the worlds the speaker can imagine, including those describing how reality could have been (but isn't). $E_a \subseteq H_a \subseteq H_0$.

Evidence for the conditions on the hypothetical conditionals are more clear in the past tense:

1. $\varphi =$"Box B is in room 1"
   (a) I know that Box B was in room 1 yesterday. ($[\![\Box P_1\varphi]\!]_{E_a,t_0} = 1$)
   (b) I know that Box B was not in room 1 yesterday. ($[\![\Box\neg P_1\varphi]\!]_{E_a,t_0} = 1$)
   (c) I don't know whether Box B was in room 1 yesterday.

[4] The distinction between these two kinds of conditionals is argued for in other places – see e.g. [Edg95].

[5] As pointed out in [Edg95], the indicative voice can actually be used in real speech for counterfactuals as well, and the subjunctive voice for epistemic conditionals. The important point here is the distinction between the two kinds of conditionals and not so much how they are expressed, so for simplicity we will stick to the dichotomy as shown in Figure 2.

($[\![\Diamond P_1\varphi \wedge \Diamond\neg P_1\varphi]\!]_{E_a,t_0} = 1$)
(d) It is completely impossible for me to imagine that Box B was in room 1 yesterday. ($[\![\Diamond P_1\varphi]\!]_{H_a,t_0} = 0$, equivalently: $H_a | P_1\varphi^{t_0} = \emptyset$)

2. If Box B had been in room 1 yesterday, then Box C [would have been / must have been / could have been] in room 2. ($P_1\varphi^{t_0} \rightsquigarrow P_1\psi^{t_0}$)

3. I know that Box B was not in room 1 yesterday. Because if Box B had been in room 1 yesterday, then Box C must have been in room 2. But we know that was not possible (according to the laws of the domain). Therefore, Box B was not in room 1.

(2) can be used after (1.b), but not after (1.a) or (1.c) (to express the idea of (1.c)+(2), the epistemic conditional should be used). The variant (3) shows how hypothetical argumentation can be used when the epistemic conditional would not be allowed (i.e. "Box B had been in room 1" is used instead of "was in room 1"). Notice that different parts of the argument use different contexts: The first, third, and fourth sentences refer to E_a, while the second to H_a. Thus this analysis does not rely on requiring a counterfactual with an impossible antecedent to be evaluated as vacuously true (as in [Lew73, p. 24]).[6]

3.3 Combining Conditionals with Tense

Figure 2 shows cases of combining the two kinds of conditionals with past, present, and future times. For the past, the epistemic conditional presupposes lack of knowledge regarding what the state of affairs actually was. For the future, the epistemic conditional also presupposes such lack of knowledge. Of course, the future is uncertain (even in our deterministic domain, it could be that the behavior of objects at time $t + 1$ is only constrained but not fully determined by their behavior at time t, in which case any of the legal possibilities might occur). However, in order to use an epistemic conditional for the future, such as "If Box B will be in room 1 tomorrow, ...", the speaker must consider it actually possible, according to his or her knowledge of the current state of affairs together with background knowledge of the laws governing the domain, that it is not completely impossible for Box B to be in room 1 tomorrow. If the speaker thinks the antecedent is completely impossible in the future, the hypothetical conditional must be used: "If Box B were to be in room 1 tomorrow, ..." (see section 4.2 for another example).

4 Analysis

The analysis so far was for a simplified domain, containing a small finite set of objects with very clear kinds of relations and interactions between them. Extending the results

[6] [And51] gave an example of a subjunctive conditional that is used to argue *for* a known case: A doctor may say "If he had taken arsenic, he would have shown just these symptoms [those which he in fact shows]", where the antecedent is not assumed to be false. This could be accounted for by relaxing the usage conditions of the hypothetical conditional (in certain situations), to make it more a "law-like" conditional that holds true over all of H_a, regardless of whether its antecedent is false in E_a.

to general natural language in other domains poses difficult challenges. A few issues and cases from the literature are reviewed and discussed.

4.1 Belief, Epistemic, and Hypothetical Contexts

Adams [Ada70] gives the following pair:

1. If Aswald didn't kill Kennedy, someone else did.
2. If Aswald hadn't killed Kennedy, someone else would have.

A speaker may agree with the first sentence while disagreeing with the second. According to the analysis here, the first sentence refers to the epistemic context. The speaker does not know whether Aswals killed Kennedy (or at least, is willing to assume for the moment this lack of knowledge), but he does know that Kennedy was killed by someone. All worlds in E_a have someone killing Kennedy, hence in all those were Aswald didn't kill Kennedy, someone else did. In contrast, the speaker of the second sentence implicates that he believes Aswald did kill Kennedy. We can model this using a belief context $B_a \subseteq E_a$, which allows the speaker to believe (in B_a) that a certain statement is true (such as "Aswald killed Kennedy") without knowing (in E_a) that it is true. The usage conditions for the hypothetical conditionals then include $[\![\neg\Diamond\varphi]\!]_{B_a} = 0$ rather than the stronger $[\![\neg\Diamond\varphi]\!]_{E_a} = 0$. The second sentence then refers to a larger hypothetical context (H_a), some of whose worlds do not have Kennedy assassinated. Thus, the same speaker can know (in E_a) that Kennedy was assassinated, believe (in B_a) that Aswals did it, not actually know that for sure (in E_a), and think (in H_a) that Aswald acted alone, and so agree with the first sentence but disagree with the second.

A parallel example for the future tense is given in [Edg95, p. 239]: We have strong evidence that one of the prisoners Smith and Jones will try to escape tomorrow, and we know that Smith is docile while Jones is adventurous. We are therefore willing to agree with the first but not the second sentence:

1. If Jones doesn't try to escape tomorrow, Smith will.
2. If Jones were not to try to escape tomorrow, Smith would.

We can model it again using B_a, E_a, and H_a: In E_a, someone escapes tomorrow, and if it's not Jones then it's Smith. Because of background knowledge about Smith and Jones, in B_a it is Jones that will escape. And in (some) H_a worlds where Jones doesn't escape, Smith doesn't either.

4.2 One or Two Conditionals?

A related issue is discussed in [Str86, p. 230], which gives the example:

> Remark made in the summer of 1964: "If Goldwater is elected, then the liberals will be dismayed".
>
> Remark made in the winter of 1964: "If Goldwater had been elected, then the liberals would have been dismayed"

Strawson comments that "the least attractive thing that one could say about the difference between these two remarks is that ... 'if ... then ...' has a different meaning in one remark from the meaning which it has in the other". However, in the second case, but not the first, it seems that the speaker implicates that he *knows* the antecedent to be false. If we change the first sentence to be "If Goldwater were to be elected in the Fall (or: If it were the case that Goldwater will be elected), then the liberals would have been been dismayed", then it would also implicate that the speaker does not believe Goldwater will be elected (see also the Jones-Smith pair in the previous section). Conversely, if we change the second sentence to "If Goldwater was elected, then the liberals were dismayed", then the implication of counterfactuality is removed, and only lack of knowledge is expressed. Therefore, the analysis of the first sentence has the structure $F\varphi \to F\psi$, whereas the second is formed $P\varphi \rightsquigarrow P\psi$.

An argument against having two kinds of conditionals is that they display the same inference patterns in natural language. This can still be explained here by the fact that their definitions have a similar structure.

4.3 Transitivity

[Edg95, p. 253] gives the following example for the failure of transitivity:

1. If (ψ) Brown had been appointed, (τ) Jones would have resigned immediately afterwards.

2. If (φ) Jones had died before the appointment was made, Brown would have been appointed.

3. If Jones had died before the appointment, Jones would have resigned immediately after the appointment.

The goal is to explain why (3), which has the form $\varphi \rightsquigarrow \tau$, does not follow from (1) $\psi \rightsquigarrow \tau$ and (2) $\varphi \rightsquigarrow \psi$.

This can be explained by identifying the background knowledge implicit in (1) and (2): For (1) it includes: $\tau \rightsquigarrow \gamma$, where $\gamma =$"Jones is alive after the appointment" (the entailment comes from the meaning of 'resigned'). The implicit knowledge for (2) includes: $\varphi \rightsquigarrow \neg\gamma$ (the knowledge comes from the meaning of 'died'). If we let our set of given assumptions be $\Gamma = \{\psi \rightsquigarrow \tau,\ \varphi \rightsquigarrow \psi\}$, and the background knowledge be $\Delta = \{\tau \rightsquigarrow \gamma,\ \varphi \rightsquigarrow \neg\gamma\}$, then we get that $\Gamma \nVdash_{\Delta} \varphi \rightsquigarrow \tau$.[7] This is because $Cxt(\Gamma \cup \Delta) = \emptyset$ (for a context C to be in $Cxt(\Gamma \cup \Delta)$, it needs to satisfy all four assumptions, which is clearly impossible). Another way of expressing the idea is saying there is a *context shift* between (1) and (2): In (1), γ is assumed, whereas in (2), $\neg\gamma$ is assumed. There is therefore no context in which both sentences can be evaluated at once (and we do not allow vacuous conclusions in such a case).

[7] Recall the definition of $\Vdash$ from section 2.2

Nevertheless, the following weakened inference pattern is still valid:

$$\frac{\psi_1 \to \psi_2 \quad (\psi_1 \wedge \psi_2) \to \psi_3}{\psi_1 \to \psi_3}$$

provided the assumptions are consistent together with the implicit knowledge.

4.4 Relativization to a Context

The central problem with counterfactuals is that, *strictly speaking*, almost all of them are false – it is almost always possible to find circumstances that provide a counterexample. This was stated in various ways in theories of conditionals, as early as [Goo47], where the sentence "If the match had been struck, it would have lit" is true only under certain qualifications about the situation (the match was dry, there was oxygen, etc.). So in what sense could hypothetical conditionals ever be true? What is the point in uttering them?

It should be pointed out that the uncertainty about the conditional is not necessarily probabilistic (as in: "If I toss this coin 100 times, it will come up heads roughly half the times"). When the speaker says (∗) "If I drop this ball, it will reach the ground in one second", he intends to express a *definite* rule – that's the whole point, to express that the laws of physics are constant, given that no "unexpected" things happen and no new factors come into play. The uncertainty is *not* of the kind: "with probability 90%, the ball will reach the ground in one second, and with probability 10% that won't happen because someone will stop it in mid air, or aliens will teleport it to their ship, or ..."

The basic way out of the dilemma is to understand that the meaning of conditionals is highly context dependent, and to relativize the meaning to particular contexts. Thus, although (∗) is strictly false because unexpected things could happen, what the speaker intends to say by uttering this sentence is that in an *idealized context*, where only the earth and the ball exist in a vacuum (plus a massless volumeless agent holding the ball, who can either let go or not at any given time point), dropping the ball would cause it to reach the ground in one second. In that idealized context, the conditional is in fact true. Theories such as [Goo47] try to start with reality and then restrict it by adding a list of qualifications. Such a list is endless, and the approach leads to circularities. Instead, constructing an idealized context, where only the minimal assumptions necessary to support the claim are stipulated, is easier and more clear. The remaining question of the faithfulness of the idealized context to actual reality is a different matter, distinct from the question of the meaning of the conditional in natural language.

The framework presented here relativizes the meaning of conditionals to a context B_a, E_a, or H_a, and to some set of background knowledge, but does not impose any structure on the context. Can we say more about the context, or where it comes from? I think this question is inherent in natural language and is not particular to conditionals, and so we may not be able to say additional general things about it. Theories of conditional, however, do try to show how some phenomena related to conditionals result from certain structural traits of the context, as discussed in the following section.

4.5 Similarity Between Worlds

A leading idea for such a structure tries to refine the set of all hypothetical worlds by using a similarity relation on them [Sta68, Lew73]. The alleged problem with a basic framework such as the one presented here[8] is that all worlds are considered equally, and so a conditional almost always comes out false. Consider: "If kangaroos had no tails, they would topple over", taken from [Lew73]. Well, that is false if you consider some of the worlds in which kangaroos with no tails evolved so that their center of gravity would be different than what it is now, allowing them to keep their balance. So how come this and similar conditionals seem true? The answer that similarity-based theories give is that if you consider only the hypothetical worlds that are as similar to the real world as possible, while still satisfying the antecedent (or in other words, those worlds that are different from the real world "just enough" to satisfy the antecedent but no more than necessary), then in those worlds the consequent is true.

The similarity relation is defined in a "once-and-for-all", objective way on all possible worlds. There are many known problems with this approach. One problematic case, discussed in [Kra89], is the following: Let w be a world such that (a) a zebra escaped; (b) it was caged with another zebra; (c) a giraffe was also in the same cage. Then the sentence "If a different animal had escaped, it might have been a giraffe" might be predicted by a similarity-based theory to be false (in w), counter intuition. The reason is that a hypothetical world which is exactly the same as w except that the other zebra had escaped is more similar to w than a world in which the giraffe had escaped. In the framework of this paper, however, the sentence is predicted to be true, since both hypothetical worlds are considered equally.

The reverse failure is noted in [NC01]: Suppose Fred's lawn is slightly too short to come into contact with the lawnmower blades, so the lawnmower currently will not cut the grass. Suppose further that the engine is weak so it will stall if the grass is to high. Then the sentence "If the grass were higher, Fred's mower would cut it" should be false, but it is predicted by similarity-based theories to be true, because worlds in which the grass is higher than the blades' height but not too high are closer to the real world than worlds in which it is too high. Again, the basic framework considers also the high-grass worlds, and says the conditional is false.

Another case, which I think is problematic, is Lewis' sequence: "If the USA threw its weapons into the sea tomorrow, there would be war; but if the USA and the other nuclear powers all threw their weapons into the sea tomorrow, there would be peace; but if they did so without sufficient precautions against polluting the world's fisheries there would be war;..." [Lew73]. Lewis' explanation considers all of these sentences to be true simultaneously,[9] based on the idea that each is "automatically" evaluated or supported by a different sphere of possible worlds, each strictly containing its predecessor. "Automatically" here means: based on the fixed similarity relation between worlds.

[8] I call the framework 'basic' because it employs possible worlds, as many theories of conditionals do, but imposes no further structure on the set of worlds (except for the distinction between contexts such as B_a, E_a, and H_a).

[9] "I think it is clear from my examples that such a sequence could consist of counterfactuals ... all of which are as definitely true as counterfactuals ever are." [Lew73, pp. 12, 18-19].

I think that no one could consider all the sentences, and in particular the first two, to be simultaneously true. There is a *shift* in the implicit context or background between successive sentences, just as in the case analyzed in section 4.3. The original context has certain stipulations that support the sentence. When moving on to the next sentence, the implicit context is changed by modifying or relaxing some of the stipulations. I doubt there is any way to define an objective definite general similarity relation that can support such shifts.

4.6 Relevance and Causality

In the zebra-giraffe case above, the basic framework correctly predicts that a might-conditional is true, whereas a similarity-based theory might predict it to be false. There are other cases where the basic theory is wrong and the other theory is correct.

Consider a puzzle that consists of two independent sub-puzzles with no interactions between the two. Suppose that in E_a, $\Box(\varphi \wedge \psi)$ is true, where φ refers to the first puzzle, and ψ to the second; and suppose that $\Diamond(\neg\varphi \wedge \neg\psi)$ in true in H_a. Then the sentence: "If φ had not been the case, then ψ might not have been the case", formalized $\neg\varphi \rightsquigarrow \Diamond\neg\psi$, is predicted to be true.[10] But in some sense it's false, because the sub-puzzles are independent, so even if φ had been false (say, as a result of the agent doing something else), this should not have influenced the truth of ψ. A concrete example is the sentence "If I had gone out of bed one minute earlier, the results of the Swedish elections might have been different" (from [Edg95, p. 257]). This is predicted by the basic theory to be true, but human intuition says this should be false because the two parts are independent.[11]

While a similarity-based theory would correctly predict the conditional to be false, because $\neg\varphi \wedge \psi$ worlds are closer to the real world than $\neg\varphi \wedge \neg\psi$ worlds, I think this prediction is based on the wrong explanation. The reason should be based on a model of *causality* between processes, expressed as a causality relation between simple propositions (a crude example: φ^t and ψ^{t+1} are causally linked iff they belong to the same sub-puzzle). A conditional $\varphi \rightsquigarrow \psi$ would be true only if there is a causal link between φ and ψ.

4.7 Point of View

Some theories define the semantics of a conditional to vary depending on the world in which it is evaluated, such that if the antecedent is true in that world, the conditional's meaning is the same as the consequent. For example, in Naive Premise Semantics [Kra81], the *would* conditional $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ is equivalent to $(\varphi \supset \psi) \wedge (\neg\varphi \supset \Box(\varphi \rightarrow \psi))$, and the *might* conditional $\varphi \mathrel{\Diamond\!\!\rightarrow} \psi$ is equivalent to $(\varphi \supset \psi) \wedge (\neg\varphi \supset \Diamond(\varphi \wedge \psi))$. Similarly, in Lewis' analysis [Lew73], in any particular world w, one can infer $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ from $\varphi \wedge \psi$. The example he gives is this:

[10] See property 2 in section 2.3.1.

[11] This problem is similar to the failure of Naive Premise Semantics, where $\varphi \mathrel{\Diamond\!\!\rightarrow} \psi$ behaves like $\Diamond(\varphi \wedge \psi)$ in case $\neg\varphi$ is true in the real world, as noted in [Kra81].

> You say: 'If Caspar had come, it would have been a good party'. I reply: 'That's false; for he did come, yet it was a rotten party.' Or else I reply: 'That's true; for he did, and it was a good party. You didn't see him because you spent the whole time in the kitchen, missing all the fun.' Either reply seems perfectly cogent.

As Lewis points out, inferring $\varphi \Box\!\!\rightarrow \psi$ from $\varphi \wedge \psi$ can be done for *any* φ and ψ that happen to be true in the world of evaluation, even if they are completely unrelated. He says that's "oddity .. not falsity".

In contrast, the framework presented here treats the conditional as a modal operator that has exactly the same truth value in all worlds (in the context of evaluation: E_a or H_a). The effect of the theories mentioned above, i.e. evaluating a conditional in a particular world w, can be achieved in our framework by taking $E_a = \{w\}$, i.e. w is known to be the real world. The difference between evaluating a conditional in a particular world vs. a context of worlds could be said to mirror the fact that the world-sensitive theories try to evaluate conditional statements "objectively", as if they make statements about objective reality, whereas the framework here relativizes the meaning of conditional statements to the contexts of knowledge and belief of a certain speaker. Moreover, the definitions above conflate the two kinds of conditionals – the epistemic and the hypothetical – into one. They insist on assigning truth values to conditional statements that in my opinion cannot be evaluated due to violation of presuppositions.

Lewis is indeed not completely sure about his definition, and discusses an alternative with a "weakly centered system of spheres", which has the effect of evaluating a conditional with a true antecedent not only in the real world but in some larger set of worlds all of which are similar to the real world according to the similarity relation. But this still conflates the two kinds of conditionals; and in any case, he abandons this idea later in his book.

5 Conclusion

The main ideas of this paper:

1. A conditional is analyzed as a modal operator w.r.t. a context of possible worlds (and has the same value in all the worlds of the context). It has a classical truth value under certain usage conditions (usually: the antecedent is possible in the context). The consequent is evaluated in a restricted context that includes those worlds in which the antecedent is true.

2. Different contexts are identified: what one believes to be true B_a, knows to be true E_a, and can imagine hypothetically H_a. The epistemic vs. hypothetical conditionals are distinguished, both in the past and the future tenses.

3. Contexts do not have internal structure, and this allows them to escape some problems of similarity-based theories. To escape other problems, they should be refined using the notion of causality.

4. A special kind of "closure" assumption – that only what was said is required – supports the non-monotonic inference for concluding what may be true given a certain set of constraints (and no others).

5. Implicit assumptions are inherent in natural language. They come into play in a background theory of assumptions, which, together with the context, may shift during a discourse or an argument, thus supporting phenomena such as the failure of transitivity, and Lewis' sequence.

References

[Ada70] Ernst W. Adams, "Subjunctive and indicative conditionals," *Foundations of Language*, vol. 6, pp. 89–94, 1970.

[And51] Alan Ross Anderson, "A note on subjunctive and counterfactual conditionals," *Analysis*, vol. 12, pp. 35–38, 1951.

[Edg95] Dorothy Edgington, "On conditionals," *Mind*, vol. 104, no. 414, pp. 235–329, 1995.

[Goo47] N. Goodman, "The problem of counterfactual conditionals," *Journal of Philosophy*, vol. 44, pp. 113–128, 1947.

[Kra81] Angelika Kratzer, "Partition and revision: The semantics of counterfactuals," *Journal of Philosophical Logic*, vol. 10, pp. 201–216, 1981.

[Kra89] Angelika Kratzer, "An investigation of the lumps of thought," *Linguistics and Philosophy*, vol. 12, pp. 607–653, 1989.

[Lew73] David Lewis, *Counterfactuals.* Harvard University Press, 1973.

[NC01] Donald Nute and Charles Cross, "Conditional logic," in *The Handbook of Philosophical Logic* (Dov Gabbay and Franz Guenthner, eds.), vol. 4, pp. 1–98, Kluwer: Dordrecht, second ed., 2001.

[Sta68] R. Stalnaker, "A theory of conditionals," in *Studies in Logical Theory*, no. 2 in American Philosophical quarterly Monograph Series, pp. 98–112, Blackwell, Oxford, 1968.

[Str86] P. F. Strawson, "'if' and '$\supset$'," in *Philosophical Grounds of Rationality* (R. E. Grandy and R. Warner, eds.), pp. 229–242, Oxford: Clarendon Press, 1986.

[Web99] Karl Weber, *The Unofficial Guide to the GRE Test.* ARCO Publishing, 2000 ed., 1999.