4 Conditionals

4.1 The material conditional

Recommended reading

James F. Thomson, "In Defense of '⊃' " (Thomson 1990). (You may skip the discussion of Strawson on pp. 61–64.)

In introductory logic, you were taught to formalize English conditionals using a truth-functional connective, the material conditional ('⊃'). You were probably also taught that this wouldn't always give good results, and that it should be considered a simplification that is useful for some purposes.

In this chapter we will ask three interrelated questions:

- Can we come up with better truth conditions for conditionals?
- If not, why not? Is it because conditionals are truth-functional after all, or because they don't have truth conditions at all?
- What logical principles hold for conditionals?

4.1.1 Indicative vs. counterfactual

It is commonly accepted that there are two fundamentally different varieties of conditionals in English (and other natural languages), *indicative* and *subjunctive*.

In English and many other languages, the distinction is marked by differences in grammatical mood. In indicative conditionals, the verb in the consequent is in the indicative mood, while in subjunctive conditions, it is in the subjunctive mood. Here is a nice minimal pair that shows the difference:

(1) If Oswald didn't shoot Kennedy, someone else did. [indicative]

(2) If Oswald hadn't shot Kennedy, someone else would have. [subjunctive]

To see the difference, ask yourself what would be evidence for each. If we think that Kennedy was, in fact, shot, then we'll accept (1) without any additional evidence. We know *someone* shot him. If it wasn't Oswald, then it must have been someone else.

By contrast, knowing that Kennedy was shot is not sufficient for accepting (2). You might accept (2) if you had evidence that Oswald didn't act alone but was part of a larger conspiracy, or if you think that so many people wanted to kill Kennedy that someone else would have stepped up. But if you think Oswald was acting alone, and an anomaly, you'd reject (2).

Because it could be rational to accept (1) and reject (2), they would seem to have different truth conditions. (2) concerns what would have happened in a possible scenario where Oswald didn't shoot Kennedy. (If in fact Oswald did shoot Kennedy, then this scenario is a counterfactual one, an alternative "possible world.") (1), by contrast, concerns what really happened in the world as it is.

You'll often hear subjunctive conditionals called *counterfactual conditionals*, because their antecedents are being presented as contrary to actual fact. But don't think that a counterfactual conditional is a conditional with a false antecedent, and an indicative a conditional with a true antecedent. A counterfactual conditional can have a true antecedent. (Suppose you mistakenly think you forget to turn in the last homework assignment, and you say: 'If I had turned in the last homework assignment, I would have passed the class'.) And an indicative conditional can a false antecedent, as in (1) above. That said, it would be weird to assert a counterfactual that you *knew* had a true antecedent, or an indicative that you *knew* had a false antecedent, and any account of the difference between indicatives and subjunctives should offer some explanation of this fact.

It seems pretty clear that subjunctive conditionals aren't material conditionals. After all, we generally use them when we know the antecedent is false. In those cases the corresponding material conditional is *always* true, but we have pretty clear intuitions that some of the subjunctive conditionals are false. Consider these pairs:

(3) a. If I were seven feet tall, I could change the light bulb. (T)
    b. If I were four feet tall, I could change the light bulb. (F)

(4) a. If I had dropped this pencil, it would have stayed on the ground. (T)
    b. If I had dropped this pencil, it would have bounced to the ceiling. (F)

When it comes to indicative conditionals, the material conditional analysis cannot so easily be dismissed. For, as noted above, we simply don't use material conditionals with antecedents that we take to be false. If I were to say

(5) If I am four feet tall, I can change the light bulb,

you would be puzzled about what I am trying to express, rather than having a clear judgment of falsity. (Am I ignorant of my own height?)

Take out a piece of paper. For each of the following sentences, write 'T' if you think it is true, 'F' if you think it is false. If you think it's wrong to call a particular sentence either true or false, you can put '?'.

(6) a. If snow is black, North Korea is in Europe.
b. If snow is black, North Korea is in Asia.
c. If snow is white, North Korea is in Asia.
d. If snow is white, North Korea is in Europe.

Think about the pattern of answers you gave, and ask others how they answered these questions. What, if anything, do these answers show?

4.1.2 Entailments between indicatives and material conditionals

One way to get clearer about the truth conditions of indicatives is to ask about entailments. In what follows, we will use the symbol '$\supset$' for the material conditional, and '$\rightarrow$' for the indicative conditional.

Nearly everyone accepts that the indicative conditional entails the material conditional:

(7) $$\frac{p \rightarrow q}{p \supset q}$$

This inference must be valid if Modus Ponens is valid for the indicative conditional. For suppose $\ulcorner p \rightarrow q \urcorner$ were true and $\ulcorner p \supset q \urcorner$ false. Then we'd have a counterexample to Modus Ponens for the indicative conditional, for the falsity of $\ulcorner p \supset q \urcorner$ requires that p be true and q false. In §4.4 we'll look at an argument that Modus Ponens in fact *fails* for the indicative conditional. But if we want our conditional to respect Modus Ponens, we'd better accept the inference (7).

More controversial is the converse entailment:

(8) $$\frac{p \supset q}{p \rightarrow q}$$

If both this and (7) were valid, we could show that $\ulcorner p \rightarrow q \urcorner$ is *equivalent* to $\ulcorner p \supset q \urcorner$, and the material conditional analysis of indicatives would be vindicated. Those who think that the indicative conditional is not a material conditional therefore reject (8). They agree that in order for $\ulcorner p \rightarrow q \urcorner$ to be true, $\ulcorner p \supset q \urcorner$ must be true, but they think that some additional connection between antecedent and consequent is required as well.

This "received opinion" is the target of Thomson 1990.

4.1.3 Thomson against the "received opinion"

Thomson makes three important observations:

- If the received opinion is correct, conditionals with false antecedents or true consequents that lack the requisite connection between antecedent and consequent are just *false*. But although we're reluctant to call them true, calling them false doesn't seem right either (Thomson 1990, p.59).
- Even if assertions of $\ulcorner p \rightarrow q \urcorner$ typically *communicate* that there is some non-truth-functional connection between p and q, it does not follow that such a connection is required for the *truth* of $\ulcorner p \rightarrow q \urcorner$.
- It may be that it is bad *reasoning* to move from $\ulcorner \neg p \urcorner$ to $\ulcorner p \rightarrow q \urcorner$. But that does not mean that $\ulcorner \neg p \urcorner$ does not *entail* $\ulcorner p \rightarrow q \urcorner$.

These last two points stand in need of more explanation.

What is said vs. what is implied

If I assert a disjunction, my audience will generally assume that I don't know which disjunct is true, since if I did, I would have made the stronger assertion. Otherwise I'd be uncooperative, and it's generally assumed that conversational partners won't withhold relevant information. For example, if I say

(9) Sam is either at the bar or studying.

you'll assume I don't know that Sam is at the bar. (This point is due to Grice 1989.)

If indicative conditionals are material conditionals, they are equivalent to disjunctions: $\ulcorner p \supset q \urcorner$ is logically equivalent to $\ulcorner \neg p \vee q \urcorner$. So

(10) If Sam is not at the bar, he is studying,

is logically equivalent to (9), and the point we just made about (9) applies here too. If I assert (10), then the normal implication is that I don't know the antecedent or the consequent. (For if I did, I'd have been in a position to make a more informative assertion, and as a cooperative conversation partner, I would have done so.) If I have reason for thinking that (10) is true, but I don't know the truth value of the antecedent or the consequent, that must be because I know something about the *relation* of the antecedent to the consequent. So, when I assert (10), others will reasonably assume that I take there to be some relation between whether Sam is at the bar and whether he is studying.

In this way, Thomson thinks, we can explain why assertions of conditionals typically communicate that there is some non-truth-functional relation between antecedent and consequent, even though the truth of the conditional doesn't require any such relation: "we read what we take to be [the speaker's] reason into the statement itself" (Thomson 1990, p. 68).

Good reasoning vs. entailment

We can all agree that inferring $\ulcorner p \to q \urcorner$ from $\ulcorner \neg p \urcorner$ looks like bad reasoning. But does it follow that $\ulcorner \neg p \urcorner$ does not *entail* $\ulcorner p \to q \urcorner$? Thomson argues that it does not. Reasoning from a premise to one of its logical consequences can sometimes be bad reasoning.

In the example at hand, this is because the conclusion is junk. If one is asserting $\ulcorner p \to q \urcorner$ solely on the basis of $\ulcorner \neg p \urcorner$, or solely on the basis of q, there is nothing one can do with the conclusion that one could not already do with the premise. Thomson illustrates this with the complex example of an oracle and an acolyte. The oracle sometimes contradicts itself, and then the acolyte has to figure out which statements to erase (Thomson 1990, p. 65). Suppose the oracle says q. It would be pointless for the acolyte to infer $\ulcorner p \to q \urcorner$, even though this follows logically. For

- If later the oracle said p, there'd be no need to do Modus Ponens to get q, because the acolyte already has q.
- If later the oracle says $\ulcorner \neg q \urcorner$, the acolyte couldn't use Modus Tollens[1] to get $\ulcorner \neg p \urcorner$. For, since his only basis for holding $\ulcorner p \to q \urcorner$ is his acceptance of q, on learning $\ulcorner \neg q \urcorner$ he would have to give up the conditional.

So there is no point to drawing the inference from q to $\ulcorner p \to q \urcorner$. That can explain our feeling that there is something wrong with this inference. But it doesn't give us grounds for thinking that the inference is invalid.

4.2 No truth conditions?

Recommended reading

Dorothy Edgington, "Do Conditionals Have Truth-Conditions?" (Edgington 1993).

Dorothy Edgington rejects the view that indicative conditionals are material conditionals, for reasons that go beyond those considered by Thomson. But, instead of proposing alternative truth conditions for indicative conditionals, she argues

[1] *Modus Tollens* is the inference from $\ulcorner p \to q \urcorner$ and $\ulcorner \neg q \urcorner$ to $\ulcorner \neg p \urcorner$.

that they do not have truth conditions at all. Conditionals, in her view, are not in the fact-stating business: they have a different role, which she seeks to explicate.

4.2.1 Arguments for the material conditional analysis

Edgington acknowledges that there are some powerful reasons for thinking that the material conditional analysis is right. Both of these inference forms seem good:

$$\text{Or-to-if}\ \frac{p \vee q}{\neg p \rightarrow q} \qquad\qquad \text{Not-and-to-if}\ \frac{\neg(p \wedge q)}{p \rightarrow \neg q}$$

Can you think of cases where you would be reluctant to make inferences with these forms?

However, if either of these inference forms are valid, $\ulcorner p \supset q \urcorner$ entails $\ulcorner p \rightarrow q \urcorner$. And we've already seen that if Modus Ponens is valid for '$\rightarrow$', $\ulcorner p \rightarrow q \urcorner$ must entail $\ulcorner p \supset q \urcorner$. So it looks like denying the equivalence of the indicative conditional and the material conditional requires either rejecting the validity of Modus Ponens, or rejecting the validity of both Or-to-if and Not-and-to-if.

Thomson has cautioned us to be skeptical about drawing conclusions about entailment from intuitions about the goodness of inferences. So there is room for maneuver here: we could try to explain why Or-to-if and Not-and-to-if inferences are good modes of reasoning *without* taking them to be valid. We'll see some examples of this strategy a bit later.

4.2.2 Arguments against the material conditional analysis

Partial acceptance

We have seen how Thomson defends the material conditional analysis against the most obvious objections by distinguishing between what is strictly speaking said and what is implied by the speaker's saying it. Edgington points out that this kind of story can at best explain why we refrain from *asserting* conditionals when we only know that their antecedents are false or their consequents true. She observes there are other data that cannot be explained in the same way: for example, data about the relation between our *degrees of confidence* in conditionals and our degrees of confidence in their antecedents and consequents.

Take an ordinary coin. What is your degree of confidence in (11)?

(11) If this coin is flipped, it will land heads.

Edgington says, plausibly, that you should have about 50% confidence in the conditional, assuming you think it's a fair coin.[2] Your degree of confidence in (11) presumably does not depend on how likely you think it is that the coin will be flipped. But the material conditional analysis predicts that it should! If the indicative conditional is a material conditional, then (11) will be true if the coin is not flipped, so you should get more confident that the conditional is true as you get less confident that the coin will be flipped. This, Edgington thinks, is a compelling reason to give up the view that indicative conditionals are material conditionals.

Thus Edgington concedes that Thomson's Gricean strategy works, if we confine ourselves to the conditionals we assert or accept with certainty. Her point is that it fails badly when we consider cases of uncertainty:

> This case against the truth-functional account cannot be made in terms of beliefs of which one is *certain*. Someone who is 100 percent certain that the Labour Party won't win has (on my account of the matter) no obvious use for an *indicative* conditional beginning 'If they win'. But someone who is, say, 90 percent certain that they won't win can have beliefs about what will be the case if they do. The truth-functional account has the immensely implausible consequence that such a person, if rational, is at least 90 per cent certain of any conditional with that antecedent. (Edgington 1993, p. 34)

Rejection

According to the material conditional analysis, *rejecting* a conditional requires accepting that its antecedent is true. But as Edgington notes, this seems wrong. For example, I might reject the conditional

(12) If the President sneezes tomorrow, the oceans will dry up.

without accepting that the President will sneeze tomorrow. I might, for example, think there's a 30% chance that the President will sneeze tomorrow, and also that there's no chance that the oceans will dry up tomorrow. In such a case, I would reject the conditional, but I would not accept the antecedent.

Bizarre validities

Edgington points out that the material implication account gives bizarre predictions about the validity of inferences. As one example, she gives William Hart's

[2] Do you agree? What is the alternative? One could simply take (11) to be false, when there is a chance that the coin will land tails, assigning the conditional a 0% degree of confidence. Against this, Edgington says: "if someone is told 'the probability is 0 that if you toss it it will land heads,' he will think it is a double-tailed or otherwise peculiar coin."

"new proof of the existence of God," which derives the existence of God from one's own failure to pray (Edgington 1993, p. 37):

1. If God does not exist, then it is not the case that if I pray my prayers will be answered. ($\neg G \rightarrow \neg(P \rightarrow A)$)

2. I do not pray. ($\neg P$)

3. Therefore (by the material conditional analysis), it is the case that if I pray my prayers will be answered. ($P \rightarrow A$)

4. So (Modus Tollens) God exists. (G)

She also notes that the material conditional analysis predicts that

$$(A \rightarrow B) \vee (\neg A \rightarrow B)$$

is a tautology. But intuitively it seems possible to reject both disjuncts. For example, if I know that Jack is on vacation in Bermuda, I might reject both 'If I go to the store, I will see Jack' and 'If I do not go to the store, I will see Jack'.

4.2.3 Rejecting Or-to-if

We noted that the inference pattern Or-to-if, together with Modus Ponens, would suffice to establish the material conditional view. Edgington does not offer a straightforward counterexample to Or-to-if: a case where we are certain of the truth of the disjunction and the falsity of the corresponding conditional. Instead, she deploys a principle linking entailment to degrees of confidence (Edgington 1993, p. 34):

(13) If A entails B, it is irrational to be more confident of A than of B.

Using this criterion, we can reject Or-to-if. Suppose you've rolled a die but you haven't seen how it landed. You think it's 1/6 likely that the die landed 1 and 1/6 likely that it landed 2. Now ask yourself:

(a) How likely is it that it landed either 1 or 2?

(b) How likely is it that, if it didn't land 1, it landed 2?

It seems rational to answer 1/3 to (a) and 1/5 to (b). But then, by our principle (13), the conditional 'if it didn't land 1, it landed 2' does not follow from the disjunction 'it landed either 1 or 2'.

If Or-to-if is invalid, why can't we find a straightforward counterexample? Edgington shows that in all *normal* cases where we would be prepared to assert a disjunction '$A \vee B$', we should have high credence in the condition '$\neg A \rightarrow B$'. By normal cases where we would assert '$A \vee B$', she means cases where

- we have intermediate credence in both disjuncts, and
- we do not accept the disjunction on the basis of one of the disjuncts alone.

As she puts it: "If I am agnostic about A, and agnostic about B, but confident that A or B, I must believe that if not-A, B" (Edgington 1993, p. 40).

However, in cases where we have low credence in the disjunction (like the die case above), and cases where we accept the disjunction only because we think one of the disjuncts is very likely, Or-to-if fails rather obviously. Suppose you're 90% confident that it's 8 o'clock, but you think there's a small chance your clock is broken. Since you're 90% confident that it's 8 o'clock, you should be at least 90% confident that

(14) It is either 8 o'clock or 11 o'clock.

(It can't be rational to be less confident in a disjunction than in one of the disjuncts.) But you don't give high credence to the conditional

(15) If it is not 8 o'clock, it is 11 o'clock.

For, if it is not 8 o'clock (because the clock is broken), it is equally likely to be any other time.

In this way, Edgington explains both why Or-to-if seems so intuitively compelling, and why it is nonetheless invalid.

4.2.4 Edgington's positive view

So, what does Edgington think are the truth conditions of indicative conditionals, if they are not material conditionals? Her view is radical. She thinks that indicative conditionals do not have truth conditions at all. Conditionals are not "part of fact-stating discourse" (Edgington 1993, p. 46).

Instead saying under what conditions conditionals are true, Edgington proposes to explain their meanings by saying what mental states they express. When we judge that if A, B, she says, we are not judging that some proposition, *that if A, B*, is true. We are, rather, judging that *B under the supposition that A*. Similarly, when we judge it 60% likely that if A, B, we are not judging that some proposition (whose truth conditions we might try to articulate) is 60% likely to be true. Rather, we are judging that B is 60% likely to be true, under the supposition that A.[3]

The core of Edgington's positive view is the principle

[3] Her approach is a kind of *expressivism*, akin to Allan Gibbard's approach to normative language (Gibbard 2003), or Huw Price's approach to the language of probability (Price 1983).

Conditional Likelihood X believes that (judges it likely that) if A, B to the extent that X judges that $A\&B$ is nearly as likely as A (Edgington 1993, p. 38).

If we represent judgments of likelihood as numerical probabilities, then this amounts to

> A person's degree of confidence in a conditional, if A, B, is the conditional probability he assigns to B given A.[4] (Edgington 1993, p. 39)

David Lewis (1976) showed that (given some plausible assumptions) there is no way to assign truth conditions to propositions of the form '$p \rightarrow q$' that will validate

The Equation

$$\Pr(p \rightarrow q) = \Pr(q|p)$$

So if Edgington is right that the degree to which you should believe $\ulcorner p \rightarrow q \urcorner$ is your subjective probability of p given q, then Lewis's triviality proof could be used as an argument for the no-truth-conditions view.[5] However, Edgington doesn't want to assume precise values, so she doesn't rely on the Lewis result. Instead she relies on an argument (discussed in the next section) that if the indicative conditional has truth conditions at all, it must be truth-functional. (She has already argued that the conditional is not truth-functional, so this suffices to establish the no-truth-conditions view.)

Because Edgington does not think that conditionals have truth values, she cannot think of validity as a matter of truth preservation. Instead, she embraces a notion of *probabilistic entailment* due to Ernest Adams.

Probabilistic validity Let the *improbability*[6] of a proposition be 1 minus its probability. An argument is *probabilistically valid* just in case the improbability of the conclusion is guaranteed to be less than or equal to the sum of the improbabilities of the premises. (That is, for every probability function, the improbabilities of the premises sum to greater than or equal to the improbability of the conclusion.)

Note that a valid argument with many premises that have a high degree of probability can have a conclusion with a low degree of probability. For an example, consider

---

[4] The *conditional probability* of B given A is defined as follows (assuming $\Pr(A) > 0$):

$$\Pr(B|A) = \frac{\Pr(A \wedge B)}{\Pr(A)}.$$

[5] See Bennett 2003, ch. 5 for an accessible exposition and analysis of Lewis's argument.

[6] Adams uses the term "uncertainty" instead, but this has the odd result that propositions we are certain are false have the highest possible "uncertainty."

(16)
The die will not land on 1. [5/6 likely]
The die will not land on 2. [5/6 likely]
The die will not land on 3. [5/6 likely]
The die will not land on 4. [5/6 likely]
The die will not land on 5. [5/6 likely]
The die will not land on 6. [5/6 likely]
———
The die will not land on 1–6. [0/6 likely]

Valid arguments will preserve certainty, but they need not preserve degree of uncertainty.

4.2.5 Against truth conditions

On Edgington's view, there is no way to assign truth conditions to an indicative conditional: "there is no proposition such that asserting *it* to be the case is equivalent to asserting that *B* is the case given the *supposition* that *A* is the case" (Edgington 1993, p. 30). We have seen why she rejects the material conditional account, which is the only plausible truth-functional account of the conditional. But as we saw in Chapter 3, it is possible to give truth conditions for non-truth-functional operators and connectives. Perhaps this can be done for the indicative conditional? To rule this out, Edgington argues that if the indicative conditional has truth conditions at all, it must be truth-functional (Edgington 1993, pp. 42–46). Since she has already argued that the indicative conditional is not truth-functional, this gives her a general argument that it lacks truth conditions.

The main premise of her argument is the Conditional Likelihood principle stated above. The argument takes the form of a "tetralemma" (like a dilemma, but with four "horns" or alternatives instead of two). Suppose truth-functionality fails. Then the truth value of a conditional is not entirely determined by the truth values of its antecedent and consequent. So at least one of the following cases must obtain:

TT Some conditionals with a true antecedent and a true consequent are true and some are false.

TF Some conditionals with a true antecedent and a false consequent are true and some are false.

FT Some conditionals with a false antecedent and a true consequent are true and some are false.

FF Some conditionals with a false antecedent and a false consequent are true and some are false.

Exercise 4.1: Material conditionals

1. ⋆Construct a slingshot argument for the conclusion that the indicative conditional is truth functional.
2. ⋆Is Edgington right that anyone who accepts truth conditions for 'if' that sometimes make a conditional with true antecedent and consequent true, and sometimes false, must accept C_1 (defined on this page)? (Would it be possible to give an account on which *certainty* that the antecedent and consequent were true would suffice for certainty in the conditional, but the mere truth of the antecedent and consequent would not suffice for the truth of the conditional?)

Case TF can be ruled out straightaway: assuming Modus Ponens is valid, a conditional with a true antecedent and a false consequent cannot be true. That leaves three cases. Edgington is going to argue that none of them is possible. That will show that truth functionality can't fail.

If Case TT can obtain, she argues, then

> C_1. Someone may be sure that *A* is true and sure that *B* is true, yet not have enough information to decide whether 'If *A*, *B*' is true; one may consistently be agnostic about the conditional while being sure that its components are true (as for '*A* before *B*').

However,

> C_1 is incompatible with our positive account [Conditional Likelihood]. Being certain that *A* and that *B*, a person must think *A*&*B* is just as likely as *A*. He is certain that *B* on the assumption that *A* is true. (Edgington 1993, p. 44)

So this possibility must be rejected. "Establishing that the antecedent and consequent are true is surely one incontrovertible way of verifying a conditional" (Edgington 1993, p. 44).

The arguments against Case FT and Case FF rely on similar reasoning. For Case FT, Edgington argues that someone who is certain that *B* will have to regard *A*&*B* as just as likely as *A*, and by Conditional Likelihood this is sufficient for being certain that if *A*, *B*. Similarly, for Case FF, Edgington argues that if someone who knows that *A* and *B* have the same truth value (as would be the case if both were false) also knows that *A*&*B* is just as likely as *A*, and hence that if *A*, then *B*.

Notice that all of these arguments move from the observation that we would be certain that if *A*, *B* if we were certain about the truth values of *A* and/or *B*, to the conclusion that it would be *true* that if A, B if A and/or B had certain truth values. Are these transitions warranted?

4.3 Stalnaker's semantics and pragmatics

Recommended reading

Robert Stalnaker, "Indicative Conditionals" (Stalnaker 1975).

Stalnaker agrees with Edgington that the material conditional analysis must be rejected, and that Or-to-if is invalid. But he gives a different sort of positive view, one that assigns truth-conditions to indicative and subjunctive conditionals in a modal framework.

4.3.1 Propositions, assertion, and the common ground

Before we look at Stalnaker's theory of conditionals, we need to sketch the theoretical background within which he gives his analysis.

The point of inquiry, as Stalnaker conceives it, is to distinguish between alternative ways the world could be. We start out in a state of ignorance. As far as we know, there are many open possibilities: the world could be this way or that way. A *possible world* is a maximally specific way the world might be: one that settles every question you could pose about the state of the world. As we inquire, we rule out possible worlds. The more opinionated we become about how things are, the fewer open possibilities remain.

A *proposition* is the content of a belief or assertion. When you believe that snow is white, for example, the thing you believe, namely *that snow is white*, is a proposition. For many purposes we can model propositions as functions from possible worlds to truth values: a proposition has the value *true* on the worlds in which it is true, and *false* on the worlds in which it is false. Equivalently, we can think of a proposition as a set of possible worlds: those at which it is true.

In this model, accepting a proposition is accepting that the actual world is a member of it. Rejecting a proposition is denying that the actual world is a member of it. And regarding a proposition as an open possibility is thinking that the actual world might be a member of it.

Assertion is a speech act that has its place in a *shared* process of inquiry. In any conversation, there is a *common ground* of propositions that are accepted within that conversation. (The participants may not actually believe these propositions, since one can accept a proposition, in the framework of a conversation, without believing it.)

The common ground is *common* in the sense that there is common knowledge about what is mutually accepted. If I am in doubt about whether you accept p, then p is not part of the common ground, even if in fact we all do accept it. Indeed, even if we all accept p, and we all *know* that the others accept p, p will fail to be in the common ground if we suspect that the others might not know that we accept p. To *presuppose* that p is to take p to be part of the common ground.

We can think of the common ground as a set of propositions. But we can also think of it as a set of possible worlds: those that fall into the intersection of the propositions. This set, which Stalnaker calls the *context set*, contains all of the worlds that are compatible with what the conversational partners mutually accept. Everything else is "off the table" and ruled out, for purposes of the conversation.

Once an assertion is made and accepted, its content is added to the common ground. We intersect the context set with the asserted proposition, throwing away worlds that aren't compatible with what is asserted. So, as the conversation progresses and more assertions are made and accepted, the context set shrinks. (Remember, removing worlds from the context set corresponds to *adding* information: the more propositions are accepted, the fewer worlds remain open possibilities.)

A proposition p is *accepted* in the context if p is true at every world in the context set.

4.3.2 Semantics

With this background in place, we can turn to Stalnaker's views about conditionals. Stalnaker thinks the truth conditions are the same for indicative and subjunctive conditionals. The difference between them has to do with the different *presuppositions* they carry.

> The idea of the analysis is this: a conditional statement, if A, then B, is an assertion that the consequent is true, not necessarily in the world as it is, but in the world as it would be if the antecedent were true. (Stalnaker 1975, p. 274)

More formally:

> $f(p, w)$ is a *selection function* that picks out the "closest" or "most similar" possible world to w at which p is true.
>
> $\ulcorner p \to q \urcorner$ is true at w if q is true at $f(p, w)$. (If $f(p, w)$ is not defined because there is no world where p is true, then the conditional is vacuously true.) (Stalnaker 1975, p. 275).

Constraints on the selection function

The selection function picks out the "closest" or "most similar" world in which the antecedent is true. But what is meant, exactly, by "closest" or "most similar"? There is no fixed answer: "Relevant respects of similarity are determined by the context" (Stalnaker 1975, p. 275). However, we can articulate three important constraints on selections functions:

C1 p is true at $f(p, w)$. (p is true at the closest world at which p is true.)

C2 If p is true at w, $f(p, w)=w$. (No world is closer to w than w itself.)

C3 If w is in the context set, then $f(p, w)$ must, if possible, be within the context set: that is, "all worlds within the context set are closer to each other than any worlds outside it."

While C1 and C2 hold for both indicatives and subjunctives, C3 is specific to indicative conditionals. When I use an indicative, the selection function has to pick out a world inside the context set. That is, the hypothetical situation we are considering must be compatible with everything we are already assuming about the actual world. When I use a subjunctive conditional, by contrast, I'm signaling that C3 does not apply: the closest world where the antecedent is true may be outside of the context set. This difference explains why subjunctives, but not indicatives, can felicitously be used when the antecedent is assumed to be false:

(17) Granted, I have a car.
But if I didn't have a car, I'd take the bus.
??But if I don't have a car, I'll take the bus.

4.3.3 Reasonable but invalid inferences

On Stalnaker's theory, the inference from q to $\ulcorner p \rightarrow q \urcorner$ is invalid. To get a countermodel, just suppose that q is true at the actual world, but false at the closest p-world.

Similarly, the inference from $\ulcorner p \vee q \urcorner$ to $\ulcorner \neg p \rightarrow q \urcorner$ (Or-to-if) is invalid. Here is a countermodel:

Context set = $\{w_1, w_2, w_3\}$

p is true at w_1 only, q is true at w_2 only.

$f(\ulcorner \neg p \urcorner, w_1) = w_3$

In this model $\ulcorner p \vee q \urcorner$ is true at w_1, but $\ulcorner \neg p \rightarrow q \urcorner$ is not true at w_1.

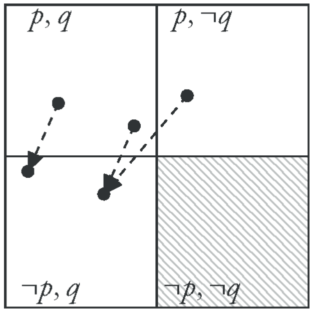


Figure 4.1: Or-to-if is a reasonable inference. The non-hatched rectangles represent the context set after $\ulcorner p \vee q \urcorner$ *has been appropriately asserted and accepted. No matter where we are in the context set, the closest world at which* $\ulcorner \neg p \urcorner$ *is true will be in the lower-left quadrant, and will therefore be a world where q is true. Thus* $\ulcorner \neg p \rightarrow q \urcorner$ *will be accepted as well. Note that the lower-left quadrant must be nonempty, since if it were empty the disjunction wouldn't have been "appropriately asserted."*

The inference from $\ulcorner \neg p \urcorner$ to $\ulcorner p \rightarrow q \urcorner$ and the inference from $\ulcorner \neg(p \wedge q) \urcorner$ to $\ulcorner p \rightarrow q \urcorner$ (Not-and-to-if) are also invalid. Of course, these inferences have to be invalid if we are to avoid the collapse of the indicative to the material conditional. But still, they can seem compelling, and we need to explain why.

Stalnaker calls an inference *reasonable* just in case, in any context where the premises "might appropriately be asserted," the conclusion will be accepted by a context if the premises are.[7] All valid inferences will be reasonable, in this sense, but some invalid inferences will also be reasonable. For an inference can be *acceptance-preserving* without being *truth-preserving*.

Or-to-if is a reasonable inference (see Fig. 4.1). Suppose $\ulcorner p \vee q \urcorner$ is appropriately asserted and accepted in the common ground C. We can now show that $\ulcorner \neg p \rightarrow q \urcorner$ must also be accepted in C. Let w be an arbitrary world in C. $\ulcorner \neg p \rightarrow q \urcorner$ is true at w just in case q is true at $w' = f(\ulcorner \neg p \urcorner, w)$. Since $\ulcorner p \vee q \urcorner$ was appropriately asserted, there must be at least one world in C at which $\ulcorner \neg p \urcorner$ is true, so by constraint C3,

[7] What is meant here by "might appropriately be asserted"? Instead of a general definition, Stalnaker offers necessary conditions that will suffice for the cases of interest to us here (Stalnaker 1975, pp. 277–8): "*It is appropriate to make an indicative conditional statement or supposition only in a context which is compatible with the antecedent. ...a disjunctive statement is appropriately made only in a context which allows either disjunct to be true without the other.*" Stalnaker clarifies the motivation for the latter condition: "If the context did not satisfy this condition, then the assertion of the disjunction would be equivalent to the assertion of one of the disjuncts alone. So the disjunctive assertion would be pointless, hence misleading, and therefore inappropriate."

$w' \in C$. Since $\ulcorner p \vee q \urcorner$ is accepted at C, $\ulcorner p \vee q \urcorner$ must be true at w'. By constraint C1, $\ulcorner \neg p \urcorner$ must be true at w'. So q must be true at w'. This suffices to show that $\ulcorner \neg p \rightarrow q \urcorner$ is true at w. Since w was an arbitrary world in C, it follows that $\ulcorner \neg p \rightarrow q \urcorner$ is accepted in C.

Thus, on Stalnaker's view, although the indicative conditional is not logically equivalent to a material conditional, it is

> ...equivalent in the following sense: in any context where either might appropriately be asserted, the one is accepted, or entailed by the context, if and only if the other is accepted, or entailed by the context. This equivalence explains the plausibility of the truth-functional analysis of indicative conditionals, but it does not justify that analysis since the two propositions coincide only in their assertion and acceptance conditions, and not in their truth-conditions. (Stalnaker 1975, p. 279)

In particular, as Stalnaker notes, the *denial* conditions for $A \vee B$ and $\neg A \rightarrow B$ are very different.

4.3.4 Contraposition and Hypothetical Syllogism

Stalnaker's semantics for conditionals makes Contraposition and Hypothetical Syllogism invalid:

Contraposition $$\frac{\phi \rightarrow \psi}{\neg\psi \rightarrow \neg\phi}$$

Hypothetical Syllogism $$\frac{\phi \rightarrow \psi \quad \psi \rightarrow \xi}{\phi \rightarrow \xi}$$

It is easy to come up with counterexamples to these forms using subjunctives. Lewis 1973, p. 35 considers the following counterexample to Contraposition:

(18)
If Boris had gone to the party, Olga would have gone.
———
If Olga had not gone, Boris would not have gone.

Let us imagine that Boris stayed away from the party solely to avoid Olga, who was there. Olga, however, would have liked the party even better had Boris been there. In this scenario, the premise is true but the conclusion false.

Stalnaker gives the following counterexample to Hypothetical Syllogism with subjunctives (from Lewis 1973, p. 33):

Exercise 4.2: Stalnaker on conditionals

1. Show that Modus Ponens is valid for Stalnaker's conditional.
2. Give a countermodel to show that Contraposition is invalid on Stalnaker's semantics.
3. Give a countermodel to show that Hypothetical Syllogisim is invalid on Stalnaker's semantics.
4. Is Contraposition with indicative conditionals a *reasonable inference*, in Stalnaker's technical sense? Either show that it is not by giving an intuitive counterexample, or prove that it is.
5. Is Hypothetical Syllogism with indicative conditionals a *reasonable inference*, in Stalnaker's technical sense? Either show that it is not by giving an intuitive counterexample, or prove that it is.

(19)
If J. Edgar Hoover [the first director of the FBI] had been born a Russian, he would have been a communist.
If he had been a communist, he would have been a traitor.
———
So, if he had been born a Russian, he would have been a traitor.

Such counterexamples are possible because the "closest" relation is not transitive. (The closest yellow house to the closest blue house to me may not be the closest yellow house to me.)

Can you think of intuitive counterexamples to these inference forms with *indicative* conditionals? If not, does that cast doubt on Stalnaker's analysis?

4.3.5 The argument for fatalism

Stalnaker gives a beautiful application of his theory to an argument for fatalism discussed by Dummett (1964). Dummett imagines a civilian reasoning as follows during an air raid:

> Either I will be killed in this raid (K) or I will not be killed. Suppose that I will. Then even if I take precautions (P) I will be killed, so any precautions I take will be ineffective (Q). But suppose I am not going to be killed. Then I won't be killed even if I neglect all precautions; so, on this assumption, no precautions are necessary to avoid being killed (R). Either way, any precautions

> I take will be either ineffective or unnecessary, and so pointless. (Stalnaker 1975, p. 280)

The civilian decides not to take shelter and is killed. Clearly something has gone wrong in this reasoning, but what?

We can formalize the argument as follows:

1	$K \vee \neg K$
2	$\quad K$
3	$\quad P \rightarrow K$
4	$\quad Q$
5	$\quad Q \vee R$
6	$\quad \neg K$
7	$\quad \neg P \rightarrow \neg K$
8	$\quad R$
9	$\quad Q \vee R$
10	$Q \vee R$

Each step is plausible. What goes wrong?

On Stalnaker's view, the problems are in lines 3 and 7. The moves from 'K' to '$P \rightarrow K$', and from '$\neg K$' to '$\neg P \rightarrow \neg K$', ares not valid arguments, but merely reasonable inferences. When 'K' is accepted at a context, '$P \rightarrow K$' is accepted too. But when we're in a subproof, our hypothetical suppositions aren't accepted at our context. Remember, our context is compatible with both 'K' and '$\neg K$'. The feature we need arguments in subproofs to have is *truth preservation*, and this one isn't truth-preserving. "So it is a confusion of validity with reasonable inference on which the force of the argument rests" (Stalnaker 1975, p. 281).

4.4 Is Modus Ponens valid?

Recommended reading

Vann McGee, "A Counterexample to Modus Ponens" (McGee 1985).

Modus Ponens is often considered a paradigm of a valid inference. In nearly all discussions of the semantics of conditionals (including the preceding three sections), it is taken for granted that Modus Ponens is valid. Vann McGee argues that this consensus is mistaken. Thinking about his argument will deepen our understanding of indicative conditionals.

4.4.1 The intuitive counterexamples

McGee begins by giving three intuitive counterexamples to Modus Ponens. In each case the major premise is a conditional with a conditional as its consequent.

The first counterexample concerns the 1980 US Presidential election, where Republican Ronald Reagan defeated Democrat Jimmy Carter. A third Republican candidate, John Anderson, ran as an independent and garnered a small fraction of the votes.

(20)
If a Republican will win the election, then if Reagan will not win, Anderson will win.
A Republican will win the election.
———
If Reagan will not win, Anderson will win.

It seems that the first premise was true, at the time of the election, because Anderson was the only other Republican candidate. And the second premise (we now know) was also true. But the conclusion was arguably false (at that context). After all, Anderson had virtually no chance of winning, and if Reagan hadn't won, Carter would have. Thinking of the conditional in the way Stalnaker recommends, the closest world to the actual world in which Reagan didn't win was a world where Carter won.

The second example concerns an animal seen from far away in a fishing net:

(21)
If that creature is a fish, then if it has lungs, it is a lungfish.
That creature is a fish.
———
If it has lungs, it is a lungfish.

The first premise is clearly true, because the only fish that have lungs are lungfish. And we may imagine that the second premise is also true (although we don't know this for sure). The conclusion, though, seems false. Lungfish are rare. If the creature in the net has lungs, it is very likely not a fish at all, but some other kind of sea animal.

The third example concerns poor Uncle Otto, who is digging a mine in his back yard, hoping to find gold or silver.

(22)

If Uncle Otto doesn't find gold, then if he strikes it rich, he will strike it rich by finding silver.
Uncle Otto won't find gold.
———
If Uncle Otto strikes it rich, he will strike it rich by finding silver.

The first premise is true, if we assume that gold and silver are the only minerals of value that could possibly be buried in the back yard. The second premise is also very likely true; it would be a huge coincidence if there were gold in Otto's back yard. But the conclusion seems false. Otto probably won't strike it rich at all, but if he does, it is just as likely to be by finding gold as by finding silver.

You might object that if we're certain of the premises of these arguments—certain, for example, that a Republican will win—then we must be certain of the conclusions (Katz 1999). Doesn't that show that the arguments are valid?

It does not. Edgington and Stalnaker have already given us several examples of inferences involving conditionals that are certainty-preserving, or acceptance-preserving, but not valid. For example:

(23) $$\frac{q}{p \to q}$$

On Edgington's account, if you are certain that q, you must be certain that if p, then q. Yet for Edgington, this inference is not valid: for, when q is not certain, it can be rational to have a lower confidence in the conditional than one has in q. On Stalnaker's account, if q is accepted in the common ground (and p is not ruled out by the common ground), $\ulcorner p \to q \urcorner$ must be accepted in the common ground too. Nonetheless, (23) is not valid. When it is not already common ground that q, it can be true that q even when the closest p-world is not a q-world.

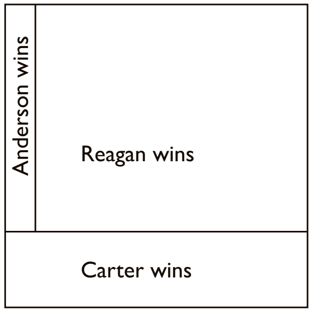
We can grant, then, that Modus Ponens is acceptance-preserving and certainty-preserving, while still raising a question about its validity. This question will have a different shape depending on whether you think of validity probabilistically (as Edgington does) or in terms of truth preservation (as Stalnaker does). So let us consider McGee's counterexamples from both points of view.

4.4.2 McGee's counterexamples as seen by Edgington

On Edgington's theory, the conclusion of a valid argument cannot have an improbability greater than the sum of the improbabilities of the premises. So, if we were to find an instance of an argument form where the two premises both have high probabilities (say, greater than 80%) and the conclusion a low probability

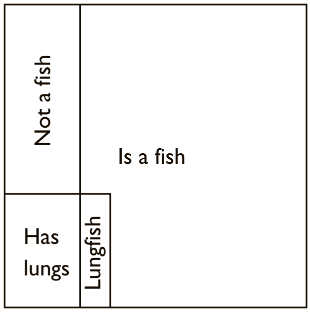
If a Republican will win the election, then if Reagan will not win, Anderson will win.
A Republican will win the election.

If Reagan will not win, Anderson will win.



If that creature is a fish, then if it has lungs, it is a lungfish.
That creature is a fish.

If it has lungs, it is a lungfish.



If Uncle Otto doesn't find gold, then if he strikes it rich, he will strike it rich by finding silver.
Uncle Otto won't find gold.

If Uncle Otto strikes it rich, he will strike it rich by finding silver.

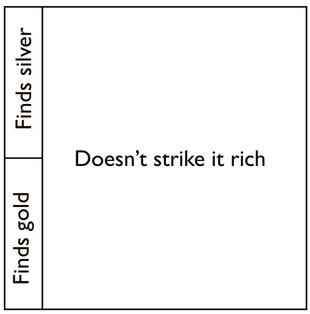


Figure 4.2: Credence diagrams for McGee's counterexamples. Bigger area = larger credence.

(say, less than 40%), that would count as a counterexample to the validity of the form.

With this in mind, examine the credence diagrams in Fig. 4.2. Here, the area occupied by a proposition represents the probability we give it. If we apply Edgington's Conditional Likelihood criterion for the acceptability of an indicative conditional, we will see that in these cases the premises have high credences, but the conclusion has a low credence. For example, in the first counterexample, the first premise presumably has a credence of 1: after we conditionalize on a Republican's winning, the conditional probability that Anderson will win given that Reagan does not win is 1. The second premise also has a high credence: it is 75% likely that a Republican will win. But the conclusion has a very low credence. Conditional on Reagan not winning, Anderson is very unlikely to win. So we have a counterexample. The other cases have different structures, but they all generate counterexamples too, given Edgington's theory of validity.

4.4.3 McGee's counterexamples as seen by Stalnaker

Let us now consider the counterexamples in Stalnaker's framework. For Stalnaker, validity is truth preservation. So, a counterexample must have true premises and a false conclusion. Of course, on Stalnaker's view, indicative conditionals are context-sensitive: they can have different truth values, and express different propositions, at different contexts. So we need to keep the context fixed in evaluating the premises and the conclusion.

Let's suppose that, shortly before the 1980 election, Sarah utters the two premises and the conclusion of McGee's first counterexample. Suppose that the context set governing her conversation at this time includes worlds where Reagan will win, worlds where Carter will win, and worlds where Anderson will win.

Clearly, the proposition expressed in this context by the second premise of McGee's argument—that a Republican would win—was true at the actual world. The actual world, as we now know, is one in which Reagan would win.

Moreover, the proposition expressed by the conclusion—that at the closest world to the actual world at which Reagan won't win, Anderson will win—is false at the actual world. Worlds where Carter wins are much more similar to actuality than worlds where Anderson wins. (An Anderson victory would have required a miracle or a stunning October surprise.)

What about the first premise? It seems awfully hard to deny that, if a Republican wins, then if it's not Reagan it's Anderson. After all, Anderson and Reagan are the only Republicans in the race. But we know that Modus Ponens is valid for Stalnaker's conditional, so this first premise *can't* be true on his account. Let's see why it isn't.

On Stalnaker's semantics, we evaluate the first premise by, first, moving to the closest world to the actual world (@) where a Republican wins—call it *w*—and then evaluating the embedded conditional at that world. But since a Republican wins at the actual world, *w* is @! So the first premise is true at @ if the embedded conditional

(24) If Reagan will not win, then Anderson will win

is true at @. And, of course, it isn't: the closest world to @ at which Reagan doesn't win is a world at which Carter wins.

Thus, Stalnaker's semantics preserves the validity of Modus Ponens, but it does so at the cost of predicting that

(25) If a Republican will win the election, then if Reagan will not win, Anderson will win.

is false (at the envisioned context). This prediction seems wrong.

One reason it seems wrong is that (25) seems equivalent to

(26) If a Republican will win the election and Reagan will not win, Anderson will win.

Interestingly, (26) *is* true on Stalnaker's semantics. The closest world at which a Republican wins and Reagan doesn't is a world where Anderson wins. So Stalnaker's semantics opens up an unexpected gap between (25), which it takes to be false, and (26), which it takes to be true. The logical rule of

$$\text{Exportation}\quad \frac{(p \wedge q) \rightarrow r}{p \rightarrow (q \rightarrow r)}$$

would allow us to infer (25) from (26). So, it seems, we have a counterexample to Exportation for Stalnaker's semantics. Saving Modus Ponens has a steep price.

4.4.4 Modus Ponens vs. Exportation

McGee shows that if we want a conditional that is stronger than the material conditional and weaker than logical implication, we need to choose between Modus Ponens and Exportation.[8] We can articulate the principle that the indicative conditional is weaker than logical implication thus:

StrImp If p logically implies q, then $\ulcorner p \rightarrow q \urcorner$ is true.

[8] The argument (McGee 1985, pp. 465–6) is similar to an argument from Gibbard 1981.

Simplifying a bit, the argument runs as follows:

1	$A \supset B$	Hyp
2	$(A \supset B) \wedge A$ logically implies B	(fact)
3	$((A \supset B) \wedge A) \rightarrow B$	StrImp 2
4	$(A \supset B) \rightarrow (A \rightarrow B)$	Exportation 3
5	$A \rightarrow B$	Modus Ponens for $\rightarrow$ 1, 4

So, if we have Exportation, Modus Ponens, and StrImp, the material conditional implies the indicative!

If we want to avoid this result, we need to give up one of these three principles. Giving up StrImp is unappealing: it means saying that a conditional could fail to be true even when the antecedent logically implies the consequent. So it's really a choice between giving up Modus Ponens and giving up Exportation. McGee argues that we should give up Modus Ponens, since there are intuitive counterexamples to Modus Ponens but not (he thinks) to Exportation. (Can you think of counterexamples to Exportation?)

Further readings

- Edgington 2014 and Bennett 2003 are useful surveys.
- Grice 1989 is an important resource for those who hope to defend the material conditional analysis of indicative conditionals. See also Rieger 2013, which summarizes a number of positive arguments for the material conditional analysis.
- On counterfactuals (not covered here), see Goodman 1955 and Lewis 1973.
- For more on the validity of Modus Ponens with indicative conditionals, see Kolodny and MacFarlane 2010.