Seminar III

Stalnaker's Theory of Conditionals

Barry Loewer and Jason Stanley

Ramsey Test: Add the antecedent of the conditional (hypothetically) to your stock of knowledge (belief), and then consider whether or not the consequent is true. Your belief about the conditional should be the same as your hypothetical belief, under this condition, about the consequent.

This clearly doesn't work for counterfactuals. So Stalnaker suggests the following as a method of evaluating any conditional, whether indicative or counterfactual:

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true.

We are skirting "the pragmatic problem of counterfactuals", namely the point that "there is more than one way to make the required adjustments".

> Now that we have found an answer to the question, 'How do we decide whether or not we believe a conditional statement?' the problem is to make the transition from belief conditions to truth conditions; that is, to find a set of truth conditions for statements having conditional form which explains why we use the method we do use to evaluate them. The concept of a *possible world* is just what we need to make this transition, since a possible world is the ontological analogue of a stock of hypothetical beliefs. (p. 33, "A Theory of Conditionals").

First approximation of Stalnaker's theory:

> Consider a possible world in which A is true, and which otherwise differs minimally from the actual world. 'If A, then B' is true (false) just in case B is true (false) in that possible world.

Model structure $\langle M,R,\lambda \rangle$. M is a set of possible worlds, R is a relation between worlds, and λ is what Stalnaker calls the *absurd world* – the world at which contradictions are true.

R is what is sometimes known as the "accessibility relation".

wRw': w' is possible relative to w

.

In addition to the standard apparatus from Kripke semantics for modal logics, Stalnaker adds to models a *selection function*:

> In addition to a model structure, our semantical apparatus includes a selection function, f, which takes a proposition and a possible world as arguments and a possible world as its value. The s-function selects, for each antecedent A, a particular possible world in which A is true. The *assertion* which the conditional makes, then, is that the consequent is true in the world selected. A conditional is true in the actual world when its consequent is true in the selected world. (p. 34)

Here is the semantical rule for the conditional:

$A>B$ is true in α if B is true in $f(A,\alpha)$
$A>B$ is false in α if B is false in $f(A,\alpha)$

$f(A,\alpha)=\beta$ – A is the *antecedent*, α is the *base world*, β is the *selected world*.

Conditions on the selection function:

(1) For all antecedents A and base worlds α, A must be true at $f(A,\alpha)$
(2) For all antecedents A and base worlds α, $f(A, \alpha)=\lambda$ only if there is no world possible with respect to α in which A is true.
(3) For all base worlds α and all antecedents A, if A is true in α, then $f(A,\alpha)=\alpha$
(4) For all base worlds α and all antecedents B and B', if B is true in $f(B',\alpha)$ and B' is true in $f(B,\alpha)$, then $f(B,\alpha)=f(B',\alpha)$.

The selection function is supposed to yield, for any antecedent A and world α, the most similar world to α at which A is true. These conditions are meant to ensure that this is what the selection function does.

The first condition ensures that the selection function picks out a world at which the antecedent is true, if it can be true. The second condition ensures that the selection function is defined even for antecedents that are impossible. The third condition ensures that the selection function is a similarity relation – nothing is more similar to α than itself. Finally, the fourth condition ensures that the similarity ordering is fixed, in the sense that we don't have a new similarity ordering when we choose different antecedents (for example, when the selection function picks out a world for A that is further out than the one it picks out for B, that is because the truth of A is a more remote possibility than the truth of B).

The selection function is a totally defined function; that is, given any values, it always yields a unique value. So Stalnaker's theory has the following consequences:

The Limit Assumption
For every possible world I and nonempty proposition A, there is at least one A-world minimally different from i.

The Uniqueness Assumption
For every world I and proposition A there is at most one A-world minimally different from i.

As we shall see, David Lewis objects to both the Limit Assumption and the Uniqueness Assumption.

Suppose that falsity is defined as failure to be true. The following principle is then validated by Stalnaker's semantics for the conditional:

Conditional Excluded Middle
(A>C) v (A>~C)

Here are some other interesting validities, which are some of the axioms of the conditional logic C2 (as in the last handout, I am using '→' for the material conditional, because, in contrast to Karen's claim, the horseshoe is NOT available on Word):

(1) □ (A → B) → (A > B)
(2) A>(B v C) → (A>B) v (A>C)
(3) (A > B) → (A→B)

As Stalnaker points out, (1) and (3) together entail that the conditional is "intermediate between strict implication and the material conditional".

It's worthwhile seeing why, e.g., (2) is valid according to the semantics. Suppose that A>(B v C) is true at w. Then f(A,w) is a world at which B v C is true. So at least one of B or C is true at f(A,w). Suppose it is B. Then A>B is true at w. Mutatis Mutandis for the supposition that C is true at f(A,w).

Now let's look at some inferences that are *invalidated* by the semantics Stalnaker provides.

Transitivity is invalid:

(1) A>B, B>C therefore A>C

Let f(A,w) be a B and ~C world, and f(B,w) be a C world.

Similarly, antecedent strengthening is invalid:

(2) A>C, therefore (A&B)>C

Let f(A,w) be a C world, and let f(A&B) be a ~C world.

'If this match were struck, it would light' does not entail 'If this match were soaked in water overnight and this match were struck, it would light'.

An inference Stalnaker does not discuss:

(3) AvC, therefore ~A>C.

This inference is invalid on Stalnaker's semantics. Suppose A is true at w, but ~C is true at f(~A,w).

We have seen (in the last handout) that the validity of (3), together with a few less problematic premises, entails that the material conditional and the conditional are logically equivalent. One of these premises is that the conditional entails the material conditional, which is valid on Stalnaker's semantics (and is in fact an axiom of C2). So the invalidity of the or-to-if inference is crucial to the success of the semantic account.

Notice that some of these inferences are fine with *indicative* conditionals:

The Direct Argument
Either the butler did it or the gardener did it. So if the butler didn't do it, the gardener did.

The inference in (a) seems perfectly in order. Indeed, it seems valid. So we need an account of why the or-to-if inference seems valid in the case of indicative conditionals, even though it is invalid according to the semantics. This is the topic of Stalnaker's paper, "Indicative Conditionals".

The direct argument seems intuitively compelling. Yet, as we have seen, if we accept it, then we must accept inferences that are clearly not intuitively compelling, as in:

Paradox of Material Implication
The butler did it. So if the butler didn't do it, the gardener did.

So here is the situation we are in. We have an inference schema that seems intuitively valid with indicative conditionals, viz. (3). But if we accept that it is valid, then we also must accept (subject to some subsidiary premises, which Stalnaker accepts) the validity of some other inference schemas that clearly do not seem intuitively valid. What to do?

Two options:

Option 1: Adopt a semantic theory for the indicative conditional that validates both inference schemas, and provide a pragmatic explanation of why one inference schema seems intuitively invalid. (Grice's strategy)

Option 2: Adopt a semantic theory that invalidates both inference schemas, and provide a pragmatic explanation of why one inference schema seems intuitively valid. (Stalnaker's strategy)

Stalnaker on the relation between the options:

> Grice's strategy and mine have this in common: both locate the source of the problem in the mistaken attempt to explain the facts about assertion and inference solely in terms of the semantic content, or truth conditions, of the propositions asserted and inferred. Both attempt to explain the facts partly in terms of the semantic analysis of the relevant notions, but partly in terms of pragmatic principles governing discourse. Both recognize that since assertion aims at more than truth, and inference at more than preserving truth, it is a mistake to reason too quickly from facts about assertion and inference to conclusions about semantic content and semantic entailment.

Stalnaker argues that although the direct argument is not valid, it is nevertheless a *reasonable inference*. Most of the paper is taken up explicating the notion of a reasonable inference.

Stalnaker on the contrast between reasonable inference and semantic entailment:

> Reasonable inference, as I shall define it, is a pragmatic relation: it relates speech acts rather than the propositions which are the contents of speech acts. Thus it contrasts with entailment which is a purely semantic relation. Here are rough informal definitions of the two notions: first, reasonable inference: an inference from a sequence of assertions or suppositions (the premises) to an assertion or hypothetical assertion (the conclusion) is *reasonable* just in case, in every context in which the premises could appropriately be asserted or supposed, it is impossible for anyone to accept the premises without committing himself to the conclusion; second, *entailment*: a set of propositions (the premises) entails a proposition (the conclusion) just in case it is impossible for the premises to be true without the conclusion being true as well.

Stalnaker gives a semantics for conditionals, together with an account of certain pragmatic features of indicative conditionals, that entails that the direct argument is a reasonable inference but not semantically valid.

Two senses of pragmatic: a Gricean sense and a Montagovian sense

Gricean sense: pragmatics fills the gap between what is said and what is communicated.

Montagovian sense: pragmatics fills the gap between the context-invariant features of sentences and what a speaker says by uttering a sentence.

For Stalnaker, which selection function is relevant for the truth-conditions of an utterance of a conditional is a context-dependent matter, and so is pragmatic in the Montagovian sense (from "A Theory of Conditionals"):

> A sentence is ambiguous if there is more than one proposition which it may properly be interpreted to express. Ambiguity may be syntactic (if the sentence has more than one grammatical structure), semantic (if one of the words has more than one meaning), or pragmatic (if the interpretation depends directly on the context of use). The first two kinds of ambiguity are perhaps more familiar, but the third kind is probably the most common in natural languages. Any sentence involving pronouns, tensed verbs, articles or quantifiers is pragmatically ambiguous. For example, the proposition expressed by 'L'etat c'est moi' deends upon who says it; 'Do it now' may be good or bad advice depending on when it is said; 'Cherchez la femme' is ambiguous since it contains a definite description, and the truth conditions for 'All's well that ends well' depends on the domain of discourse. If the theory presented above is correct, then we may add conditional sentences to this list. The truth-conditions for 'If wishes were horses, then beggars would ride' depend on the specification of an s-function.

We need to develop some elements from the theory of context-dependence to get at Stalnaker's account of reasonable inference.

We take a proposition to be a set of possible worlds. A proposition is true at a world if and only if that world is a member of the proposition.

A speaker's context set: her presumed common ground; the set of propositions she believes to be commonly presupposed in the conversational context.

We can take a speaker's context set to be a set of worlds; the intersection of the propositions that she believes to be commonly presupposed in the conversational context.

A defective context: one in which different participants in the conversation have different context sets. We're only going to be speaking of non-defective contexts. For this reason, we're just going to be speaking about the context set, rather than individual speakers' context sets.

A proposition is *compatible* with the context set if and only if there is at least one world in the context set in which it is true. A proposition is *entailed* by the context set if and only if it is true in all worlds in the context set.

Intuitively: a proposition is compatible with the context set if and only if it is an open epistemic possibility for conversational participants; it is entailed by the context set if and only if it is presupposed.

Pragmatic contextual constraint on selection functions:

> …if the conditional is being evaluated at a world in the context set, then the world selected must, if possible, be within the context set as well (where C is the context set, if i∊C, then f(A,i)∊C). In other words, all worlds within the context set are closer to each other than any worlds outside it.

Stalnaker on the motivation for the 'contextual constraint':

> The motivation of the principle is this: normally a speaker is concerned only with possible worlds within the context set, since this set is defined as the set of possible worlds among which the speaker wishes to distinguish. So it is at least a normal expectation that the selection function should turn first to these worlds before considering *counterfactual* worlds—those presupposed to be non-actual.

So the idea is that this constraint is one that flows from general conversational principles. In a conversation, one is only concerned with communicating information about live conversational possibilities (or else one would violate Grice's maxim of Relation). So, in uttering a conditional proposition, one wants to distinguish between the live conversational possibilities. The contextual constraint follows from these facts about the purpose of communication.

What about counterfactual conditionals? The treatment of counterfactual conditionals flows from Stalnaker's claims about the subjunctive mood. This is, for Stalnaker, a conventional feature of the subjunctive mood:

> I take it that the subjunctive mood in English and some other languages is a conventional device for indicating that presuppositions are being suspended, which means in the case of subjunctive *conditional* statements, that the selection function is one that may reach outside of the context set. Given this conventional device, I would expect that the pragmatic principle stated above should hold without exception for indicative conditionals.

That is, given that natural languages have a special device for indicating that presuppositions are being suspended (namely the subjunctive mood), it follows that we should not expect to find uses of indicative conditionals that are not in accord with the pragmatic principle. For if the speaker wished to distinguish among possibilities that are not in the context set, she would employ the subjunctive mood.

(Think about "will" conditionals. Does Stalnaker's claim about the subjunctive mood entail that "will" conditionals are not subjunctive, contra Dudman?)

Propositions assumed to be false are of course not compatible with the context set. So, if one wishes to speak about a counterfactual antecedent, one must employ the subjunctive mood. So, one will intend a selection function that reaches outside the context set. So utterances of counterfactual conditionals are associated with different selection functions than utterances of indicative conditionals.

As Stalnaker points out, the fact that all indicative conditionals conform to the pragmatic constraint on the selection function discussed above has the following consequence for the *appropriateness conditions* of uttering indicative conditionals:

> It is appropriate to make an indicative conditional statement or supposition only in a context which is compatible with the antecedent.

That is, it is appropriate to make an indicative conditional statement only if the truth-value of the antecedent is unknown. This is a widely accepted claim about the appropriateness conditions for indicative conditionals (see the Strawson reading).

Before turning to the direct argument, Stalnaker adds a further claim about appropriateness, this time about the appropriateness of disjunctive statements:

> A disjunctive statement is appropriately made only in a context which allows either disjunct to be true without the other. That is to say, one may say A or B only in a situation in which both (A and ~B) and (B and ~A) are open possibilities.

Stalnaker argues that this follows from general principles about conversation. Suppose that (A and ~B) is not an open possibility. Then, it is known that (~A or B). The effect of uttering 'A or B' would then be the very same as the effect of uttering B alone, since ~A or B together with A or B entails B. So adding the disjunct would be conversationally pointless. *Mutatis Mutandis* for (B and ~A).

Now we have the resources to show why the direct argument is a reasonable inference. We need to show that in any context in which 'P v Q' is assertible and accepted, '~P>Q' must be accepted. That is, we need to show that, if PvQ is assertible, and subsequently asserted, then the resulting context set entails ~P>Q.

PvQ
~P>Q

Suppose that P v Q is assertible. Then, by the appropriateness condition for disjunctions, ~P and Q is compatible with the context (i.e. there is at least one world in the context in which it is true). Now suppose that PvQ is accepted. The result is a context set with some (~P&Q) worlds, but no (~P & ~Q) worlds. Since there are no (~P&~Q) worlds left, and some (~P&Q) worlds, every world will be one in which ~P>Q, i.e. ~P>Q will be entailed by the resulting context set.

This means (via the direct argument for reasonable inference) that the inference from a material conditional to an indicative conditional, and vice-versa, is always a reasonable one. "This equivalence explains the plausibility of the truth-functional analysis of indicative conditionals, but does not justify that analysis since the two propositions coincide only in their assertion and acceptance conditions".

Some possible confusions: If the antecedent is true, then the conditional is always evaluated with respect to the actual world, even if the actual world is not in the context set (or else Modus Ponens would be invalid).

(Assuming no Kenyans are running)

(1) If John ran the marathon in under three hours, he won.
(2) John ran the marathon in under three hours.
(3) John won.

Suppose that in fact some Kenyans ran the race, contra our assumptions. Then the actual world is not in the context set. If we only evaluated indicative conditionals within the context set, and Modus Ponens were valid, then we could draw the conclusion that John won, even though he lost. So the actual world is always the most similar. We are to draw on worlds on the context set only where the conditional is being evaluated at a world in the context set. If the actual world is not within the context set (because, say, we have relevant false beliefs), then it will not be possible for the selection function to select a world in the context set, if the antecedent is true (by property (3) of selection functions).