Semantics & Pragmatics Volume 11, Article 8, 2018
https://doi.org/10.3765/sp.11.8

This is an EARLY ACCESS version of

Mandelkern, Matthew. 2018. The case of the missing 'If': Accessibility relations in Stalnaker's theory of conditionals. *Semantics and Pragmatics* 11(8). https://doi.org/10.3765/sp.11.8.

This version will be replaced with the final typeset version in due course. Note that page numbers will change, so cite with caution.

The case of the missing ‘If’: Accessibility relations in Stalnaker’s theory of conditionals[*]

Matthew Mandelkern
All Souls College, Oxford

Abstract A part of Stalnaker (1968)’s influential theory of conditionals has been neglected, namely the role for an accessibility relation between worlds. I argue that the accessibility relation does not play the role intended for it in the theory as stated, and propose a minimal revision which solves the problem, and brings the theory in line with the formulation in Stalnaker & Thomason 1970.

Keywords: conditionals, accessibility relations, (reverse) Sobel sequences, indicative versus subjunctive conditionals

1 Accessibility in Stalnaker 1968

A part of Stalnaker (1968)’s influential theory of conditionals has been neglected, namely the role for an accessibility relation between worlds. I believe this is because the accessibility relation does not play the role intended for it in the theory as stated. I propose a minimal revision which solves the problem, and brings the theory in line with the formulation in Stalnaker & Thomason 1970.

In Stalnaker’s theory, a conditional $A > B$ is evaluated relative to a *model structure* and a *selection function*.[1] A model structure $\mathfrak{M}$ is a triple $\langle K, R, \lambda \rangle$, where K is a set of possible worlds; R is a reflexive binary accessibility relation on K, with $\alpha R \beta$ read ‘β is possible with respect to α’; and λ is the absurd world where every sentence is true, accessible from no other world under R. A selection function f is a function which takes a sentence and a world as its arguments, and returns a world as its value. Relative to a background model $\mathfrak{M}$ and accessibility relation f, $A > B$ is true in α iff B is true in $f(A, \alpha)$. Stalnaker imposes four conditions on f (the names are later additions):

(1) *Success*: For all antecedents A and base worlds α, A must be true in $f(A, \alpha)$.

* Thanks to two anonymous referees for this journal and my editor, Thony Gillies, and to Daniel Rothschild and Robert Stalnaker, for very helpful discussion.

1 I omit corner quotes for readability.

(2) *Absurdity*: 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) *Strong Centering*: For all antecedents A and base worlds α, if A is true in α, then $f(A,\alpha)=\alpha$.

(4) *CSO*: 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 accessibility relation R figures only in the second condition, *Absurdity*. This condition does not seem to be strong enough, however. Later in the paper, Stalnaker proposes that we can define the ordinary modal operators $\Box$ and $\Diamond$ out of his conditional semantics. In particular, he proposes this definition for $\Box$:

Defined Semantics: $\Box A =_{DF} \neg A > A$

Now, if $\Box$ is meant to receive its ordinary semantic interpretation, then we should also have the following truth conditions, relative to a model structure $\langle K,R,\lambda\rangle$:

Standard Semantics: $\Box A$ is true at α iff $\forall\beta:\ \alpha R\beta \to A$ is true at β

Stalnaker is not explicit that he intends $\Box$ to receive the *Standard Semantics*, but he uses $\Box$ in an altogether standard way (for instance, in defining the strict conditional). So we should expect *Standard Semantics* to match *Defined Semantics*.

But it doesn't: if $\Box A$ is true according to the *Defined Semantics*, then it is true according to the *Standard Semantics*, but not *vice versa*. Suppose that $\Box A$ is true at α on *Defined Semantics*. Then $\neg A > A$ is true at α, which can only hold if $f(\neg A,\alpha)=\lambda$, which, according to *Absurdity*, only holds if all the worlds accessible from α make $\neg A$ false, which guarantees that all the worlds accessible from α make A true, so $\Box A$ is true according to *Standard Semantics*. But now suppose that $\Box A$ is true on *Standard Semantics*. Then every world accessible from α makes A true. It is consistent with all the conditions above, however, that $f(\neg A,\alpha)$ is some world $\beta\neq\lambda$ which makes $\neg A$ true. Given the assumption that all the worlds accessible from α make A true, β is not accessible from α under R; but nothing rules out a situation in which β is nonetheless selected as the closest $\neg A$-world. In other words, given the conditions as stated, *inaccessible, non-absurd worlds can be selected*. In that case, $\neg A > A$ will *not* be true at α, and so according to the *Defined Semantics*, $\Box A$ will be false. Thus *Defined Semantics* comes apart from *Standard Semantics*.

For a simple model of this divergence, let $K=\{\alpha,\beta,\lambda\}$, with A true at α and false at β. Let $R=\{\langle\alpha,\alpha\rangle,\langle\beta,\beta\rangle,\langle\lambda,\lambda\rangle\}$. Let $f(\neg A,\alpha)=\beta$. This model is clearly consistent with all the constraints above. According to *Standard Semantics*, $\Box A$ is true at α, since α can only access A-worlds. But according to *Defined Semantics*, $\Box A$ is false at α, since $\neg A > A$ is false at α, since A is false at $f(\neg A, \alpha) = \beta$.

2 The missing 'If'

The accessibility relation thus seems not to be doing enough in Stalnaker's semantics. The proper role for the accessibility relation seems to me to be the following:

> *Accessibility*: For all antecedents A and worlds α, if there is some world β such that $\alpha R \beta$ and β makes A true, then α must access $f(A, \alpha)$ under R; otherwise $f(A, \alpha) = \lambda$.

That is, f should always take us to an accessible world, if it's possible to do so while satisfying *Success*, and otherwise must take us to λ. *Accessibility* does not follow from Stalnaker's conditions: the toy model just given is a countermodel, since α does not access any $\neg A$-world under R, but $f(\neg A, \alpha) \neq \lambda$.

There are two reasons to think that *Accessibility* gives the accessibility relation its proper role in Stalnaker's semantics. First, *Accessibility* is exactly what we need to render *Defined Semantics* equivalent to *Standard Semantics*.[2] Second, *Accessibility* is entailed by the semantics given in Stalnaker & Thomason 1970, which is a formal companion to Stalnaker 1968 whose goal is to extend the formal treatment of the theory given in Stalnaker 1968. In the propositional fragment of Stalnaker & Thomason (1970)'s system, the four conditions on selection functions and the semantic clauses are all equivalent to those in Stalnaker 1968. There is one crucial difference, however: in the definition of selection functions, Stalnaker & Thomason (1970) stipulate that for all A and α, if $f(A, \alpha) \neq \lambda$, then $\alpha R f(A, \alpha)$. I'll refer to this stipulation by its heading in Stalnaker & Thomason 1970, namely *D3.4*.

D3.4 entails *Accessibility*, given Stalnaker's other conditions. Actually, it is strictly more than we need: all that we need is the stipulation that, if $f(A, \alpha) \neq \lambda$,

2 *Accessibility* obviously renders these equivalent. To see that the failure of *Accessibility* lets these diverge, suppose *Accessibility* is false. Then for some antecedent A and world α, *either* there is some world β such that $\alpha R \beta$ and β makes A true, $f(A, \alpha) = \gamma$ for some world γ, and $\neg \alpha R \gamma$. Note first that, by *Absurdity*, $\gamma \neq \lambda$, since there is an accessible A world from α. Now let $\{\gamma\}$ be the proposition true just at γ and λ and false everywhere else in K (for convenience, we also use $\{\gamma\}$ to denote the sentence true just at γ and λ and false everywhere else in K). By *Success*, $f(\{\gamma\}, \alpha)$ is either γ or λ. But it can't be λ. For since A is true in λ and $\{\gamma\}$ true in λ, by *CSO* it would follow that $\gamma = \lambda$, contrary to what we have shown. So $f(\{\gamma\}, \alpha) = \gamma$. It follows that $\{\gamma\} > \neg\{\gamma\}$ is false at α, and so $\Box\neg\{\gamma\}$ is false at α according to *Defined Semantics*. But now note that since $\neg \alpha R \gamma$ and $\neg \alpha R \lambda$, we have $\Box\neg\{\gamma\}$ true at α according to *Standard Semantics*. *Or* there is no world β such that $\alpha R \beta$ and β makes A true, and $f(A, \alpha) = \gamma \neq \lambda$. By *Success*, A is true in γ, and since $\gamma \neq \lambda$, $\neg A$ is thus false in γ, so $A > \neg A$ is false in α. Thus $\Box\neg A$ is false in α according to *Defined Semantics*; but it is true in α according to *Standard Semantics*.

then there is a world possible with respect to α in which A is true (this follows from *D3.4* given *Success*, but not *vice versa*). Note that this, in turn, is just the converse of *Absurdity*. And so all we need to get to *Accessibility* from Stalnaker (1968)'s formulation is an extra 'if' in *Absurdity*, as follows:

> *Absurdity Biconditional*: For all antecedents A and base worlds α, $f(A,\alpha) = \lambda$ *if and only if* there is no world possible with respect to α in which A is true.

Given Stalnaker's other conditions, *Absurdity Biconditional* entails *Accessibility*;[3] and *D3.4* follows immediately from *Accessibility*.

Thus we should replace *Absurdity* with *Absurdity Biconditional*. This gives the accessibility relation the role that it seemed destined for, and it brings Stalnaker (1968)'s informal exposition in line with Stalnaker & Thomason (1970)'s formal development. This has two upshots, both of them small but worth noting. First, we should treat Stalnaker & Thomason (1970)'s exposition, rather than Stalnaker (1968)'s, as canonical. Second, in the subsequent literature, Stalnaker's accessibility relation has been largely ignored (perhaps partly because of the limited role that it played in the theory as stated). But in more recent literature, it has been recognized that an accessibility relation of some kind may play an important role in the theory of conditionals. For instance, an accessibility relation may play a role in distinguishing indicative from subjunctive conditionals (Stalnaker 1975, von Fintel 1998); and an accessibility relation may play a role in accounting for reverse Sobel sequences (von Fintel 2001, Gillies 2007, Williams 2008, Moss 2012). The idea that the interpretation of conditionals depends not only on a selection function of some sort but also on an accessibility relation was already present in Stalnaker 1968 — albeit missing an 'if'.

References

von Fintel, Kai. 1998. The presupposition of subjunctive conditionals. In Orin Percus & Uli Sauerland (eds.), *The interpretive tract*, vol. 25, 29–44. Cambridge,

3 Suppose first that there is some world β such that $\alpha R \beta$ and β makes A true. Suppose next for *reductio* that $f(A,\alpha) = \gamma$ and $\neg \alpha R \gamma$. We know by the left-to-right direction of *Absurdity Biconditional* that $\gamma \neq \lambda$. Now consider what $f(\{\gamma\},\alpha)$ is (where $\{\gamma\}$ again stands for the sentence true at γ and λ and false everywhere else in K). Since by assumption $\neg \alpha R \gamma$, the right-to-left direction of *Absurdity Biconditional* entails that $f(\{\gamma\},\alpha) = \lambda$. But now note that A is true at the closest $\{\gamma\}$-world, λ; and $\{\gamma\}$ is true at the closest A world, namely γ; and so by *CSO*, γ must be λ, contrary to our assumption. So our assumption for *reductio* is discharged, and we conclude that $\alpha R \gamma$, as required by *Accessibility*. Suppose next that there is no world β such that $\alpha R \beta$ and β makes A true. Then the right-to-left direction of *Absurdity Biconditional* immediately entails that $f(A,\alpha) = \lambda$, as desired.

Massachusetts: MIT Working Papers in Linguistics. https://philpapers.org/rec/VONTPO-2.

von Fintel, Kai. 2001. Counterfactuals in a dynamic context. In *Ken Hale: A life in language*, 123–152. MIT Press. http://web.mit.edu/fintel/fintel-2001-counterfactuals.pdf.

Gillies, Anthony. 2007. Counterfactual scorekeeping. *Linguistics and Philosophy* 30(3). 329–360. https://doi.org/10.1007/s10988-007-9018-6.

Moss, Sarah. 2012. On the pragmatics of counterfactuals. *Noûs* 46(3). 561–86. https://doi.org/10.1111/j.1468-0068.2010.00798.x.

Stalnaker, Robert. 1968. A theory of conditionals. In Nicholas Rescher (ed.), *Studies in logical theory*, 98–112. Oxford: Blackwell. https://philpapers.org/rec/STAATO-5.

Stalnaker, Robert. 1975. Indicative conditionals. *Philosophia* 5(3). 269–86. https://philpapers.org/rec/STAIC.

Stalnaker, Robert C. & Richmond H. Thomason. 1970. A semantic analysis of conditional logic. *Theoria* 36(1). 23–42. https://doi.org/10.1111/j.1755-2567.1970.tb00408.x.

Williams, J. Robert G. 2008. Conversation and conditionals. *Philosophical Studies* 138(2). 211–223. https://doi.org/10.1007/s11098-006-9032-5.

Matthew Mandelkern
All Souls College
University of Oxford
OX1 4AL
United Kingdom
matthew.mandelkern@all-souls.ox.ac.uk