Erkenntnis (2021) 86:59–70
https://doi.org/10.1007/s10670-018-0093-6

ORIGINAL RESEARCH

A Stalnaker Semantics for McGee Conditionals

Kurt Norlin[1]

Received: 11 August 2018 / Accepted: 5 December 2018 / Published online: 7 February 2019

Abstract
The semantics Vann McGee gives for his 1989 conditional logic is based on Stalnaker's 1968 semantics but replaces the familiar concept of truth at a world with the novel concept of truth under a hypothesis. Developed here is a semantics of the standard type, in which sentences are true at worlds, only with additional constraints imposed on the accessibility relation and the selection function. McGee conditionals of the form $A \Rightarrow X$ are translated into Stalnaker conditionals of the form $\Box A > X$. An interpretation of the semantics is provided, and a few implications for the theory of indicative conditionals and their probabilities are noted.

1 Introduction

McGee (1989) describes a conditional logic that weakens modus ponens in order to uphold the Import–Export principle, by which "If *X*, then if *Y*, then *Z*" is equivalent to "If *X* and *Y*, then *Z*." The semantics McGee gives for his system is based on Stalnaker's (1968) but replaces the familiar concept of truth at a world with the novel concept of truth under a hypothesis. McGee's semantics is not really, in other words, a standard Stalnaker semantics.

However, it is possible to provide a formally orthodox Stalnaker semantics, in which sentences are true at worlds in the usual way, and truth under hypotheses plays no role. That is what is done here, by subjecting Stalnaker's models to additional constraints, thereby shrinking the class of permitted models, and then writing McGee's $\Rightarrow$ connective in terms of Stalnaker's $>$ connective. Following McGee, $\Rightarrow$ represents the indicative if–then, rather than, indifferently, either the indicative or the subjunctive/counterfactual form, as in McGee's earlier (1985) treatment.

A brief concluding discussion explains how the new semantics is obtained and offers an epistemic or doxastic interpretation of it, and then briefly notes some ways in which the present investigation contributes to our understanding of indicative conditionals: it exposes the root cause of McGee-type modus ponens failure, it turns

✉ Kurt Norlin
k.norlin@dslextreme.com

[1] Bittner Development Group, Irvine and Redwood City, CA, USA

up evidence that the Import–Export principle is itself not immune to failure, and it provides a way to interpret the ⇒ connective as what Lewis (1976) calls a probability-revision conditional.

2 Definitions

We begin with three sets of definitions.

2.1 Stalnaker's Logic and Models

Here we recapitulate Stalnaker's syntax and model theory, with the intent of preserving the essentials of Stalnaker's original formulation, while departing from it in some incidentals in order to facilitate what follows.

A *Stalnaker sentence* is a sentence built up from atomic sentences (p, q, …) using the familiar Boolean connectives $\neg$ and $\wedge$ and the conditional connective $>$. The connectives $\vee$, $\supset$ and $\equiv$ are defined via $\neg$ and $\wedge$ in the usual way. Metalogical variables for sentences are X, Y and Z. The standard necessity operator, $\Box$, is defined by $\Box X =_{\text{df}} \neg X > X$.

A *Stalnaker frame*, $\langle W, R, f \rangle$, consists of a set W, a relation R, and a function f:

- W is a nonempty set of possible worlds.
- R is a binary relation between worlds in W. When xRy, we say that world y is *accessible* from world x. R is reflexive.
- f is a two-place function: where S is a subset of W and w is a member of W, $f(S, w)$ picks out a member of S accessible from w and is undefined just in case S contains no such member.[1] If w itself is in S, then $f(S, w) = w$. Also, if subset S_2 of S_1 contains $f(S_1, w)$, then $f(S_2, w) = f(S_1, w)$. $f(S, w)$ is to be read as denoting, of all the members of S accessible from w, that member *most similar*, or *closest* (in resemblance), to w.

A *Stalnaker model* is a Stalnaker frame plus a function I that maps each atomic Stalnaker sentence to a subset of W. Given I, the function V_S maps any Stalnaker sentence X to the set of *worlds where X is true*—or, more neutrally, 'worlds' where X is 'true,' with the scare-quoted terms standing in for concepts to be explained when we turn from the formal properties of the semantics to the question of interpretation. V_S is defined recursively:

- For atomic X, $V_S(X) = I(X)$.
- $V_S(\neg X)$ is the W-complement of $V_S(X)$.
- $V_S(X \wedge Y)$ is the intersection of $V_S(X)$ and $V_S(Y)$.
- $V_S(X > Y)$ is the set of all w such that $f(V_S(X), w)$ is in $V_S(Y)$ or else is undefined.

[1] The requirement that $f(S, w)$ be accessible from w is not properly built into Stalnaker's original exposition. This error is corrected in Stalnaker and Thomason's (1970) formulation; see Mandelkern (2018a).

- In consequence of the preceding, $V_S(\Box X) = V_S(\neg X > X)$ is the set of all w such that $f(V_S(\neg X), w)$ is undefined, i.e., the set of all w such that x is in $V_S(X)$ whenever wRx.

For convenience, we may write $f(X, w)$ to mean $f(V_S(X), w)$.

A Stalnaker sentence X is *Stalnaker-valid* just in case $V_S(X) = W$ for every Stalnaker model.

2.2 McGee's Logic and Models

The second set of definitions summarizes the elements of McGee's syntax and model theory. Again, the intent is to keep the essentials of McGee's version intact while making changes in formulation for the sake of the present purpose.

A *McGee sentence* is a sentence built up from atomic sentences (p, q, …) using $\neg$, $\wedge$, and the conditional connective $\Rightarrow$, where $\Rightarrow$ cannot occur within the antecedent of another $\Rightarrow$. The connectives $\vee$, $\supset$ and $\equiv$ are defined via $\neg$ and $\wedge$ in the usual way. The metalogical variables for sentences are A, B, C, X, Y and Z, where X, Y and Z can be sentences of any form, but A, B and C must be Boolean (i.e., free of $\Rightarrow$).

A *McGee frame*, $\langle W, R, f, w_0 \rangle$, is a Stalnaker frame with one member of W, w_0, designated as the actual world. Relation R is such that every world is accessible from w_0.

A *McGee model* is a McGee frame plus a function, I, that maps each atomic McGee sentence to a subset of W. Given I, the function V_M maps a McGee sentence X to the set of all *hypotheses under which X is true*, where a hypothesis is a subset of W. (So whereas I maps into $P(W)$, the power set of W, V_M maps into $P(P(W))$.) The empty set is in $V_M(X)$ for every X. The remaining membership of $V_M(X)$, consisting of nonempty subsets of W, is defined recursively:

- For atomic X, S is in $V_M(X)$ iff $f(S, w_0)$ is in $I(X)$.
- S is in $V_M(\neg X)$ iff it is not in $V_M(X)$.
- S is in $V_M(X \wedge Y)$ iff it is in both $V_M(X)$ and $V_M(Y)$.
- S is in $V_M(A \Rightarrow Y)$ iff $S \cap V_S(A)$ is in $V_M(Y)$, where $V_S(A)$ is the set of worlds at which A is true in the Stalnaker sense. (Since A contains only Boolean connectives, it is both a McGee sentence and a Stalnaker sentence.)

A McGee sentence X is *McGee-valid* just in case W is in $V_M(X)$ for every McGee model.

2.3 "Stalnaker-Plus" Models, and McGee-to-Stalnaker Translation

The last set of definitions describes some additional constraints imposed on Stalnaker models to obtain a smaller class of models, and specifies how to translate a McGee sentence into a corresponding Stalnaker sentence.

A *Stalnaker⁺ model* is a Stalnaker model that meets the following additional criteria:

- *Transitivity of R*. For any worlds w, x and y: if wRx and xRy, then wRy.
- *Terminality of R*. For any world w: there is a world z such that wRz, and zRx only for $x = z$. Such a z is called a *terminal world*.
- *Fusion under R*. For any worlds w, x and y: if wRx and wRy, then there is a unique world v such that wRv, vRx, vRy, and for any terminal world z, vRz only if either xRz or yRz. This world v is called the *fusion of x and y with respect to w*. The name "fusion" signals resemblance to a principle in mereology, according to which for any two objects a and b that are parts of a larger object c, there exists a unique part of c that contains only what is necessary to contain a and b.
- *Iteration of f*. For any world w and any subsets S_1 and S_2 of W such that S_1 is forward-closed under R (if w is in S_1 and wRx, then x is in S_1) and closed under all fusions (if w and x are in S_1 and v is the fusion of w and x with respect to some y, then v is in S_1): $f(S_2, f(S_1, w)) = f(S_1 \cap S_2, w)$, unless both sides of the identity are undefined.
- *Agreement under I*. Let T be the set of terminal worlds. For any world w and any atomic sentence X: w is in $I(X)$ just in case $f(T, w)$ is in $I(X)$. That is, every world agrees with its closest terminal world about the truthvalues of all atomic sentences. From agreement under I, it follows very directly by induction that for any w and any Boolean sentence X, w is in $V_S(X)$ just in case $f(T, w)$ is.
- Transitivity and fusion imply that R is antisymmetric: if wRx and xRw, then $x = w$. Proof: Suppose that wRx and xRw. By transitivity (and even without reflexivity) wRw, so that there is a unique v that is the fusion of w and x with respect to w. Since xRx, as well, w and x both qualify as v with regard to the requirement that wRv, vRw, and vRx. Also, for any terminal world z, by transitivity wRx and xRw together imply that wRz iff xRz, so that wRz iff either wRz or xRz, and similarly xRz iff either wRz or xRz. So w and x each qualify as a fusion of w and x with respect to w. But the fusion is unique, so $x = w$.
- By implying antisymmetry, transitivity and fusion also imply that f is *monotonic*, meaning that for any worlds w, x and y such that wRx and xRy: $f(\{x, y\}, w) = x$. Proof: Given that by transitivity wRy, suppose $f(\{x, y\}, w) = y \neq x$. By antisymmetry, it is not the case that yRx, and so $f(\{x\}, f(\{x, y\}, w)$ is undefined even though $f(\{x\}, w)$ is defined. But $f(S_2, f(S_1, w))$ is defined just in case $f(S_1 \cap S_2, w)$; this follows from iteration but also can be shown without it. Therefore, by reductio, $f(\{x, y\}, w) = x$. The descriptor "monotonic" conveys the fact that an R-step away from a world w can produce a drop in similarity to w but never a rise.

A Stalnaker sentence is *Stalnaker⁺-valid* just in case $V_S(X) = W$ for every Stalnaker⁺ model.

Finally, a *translation* of a McGee sentence is a Stalnaker sentence obtained from the McGee sentence by replacing every instance of $X \Rightarrow Y$ with $\Box X > Y$.

3 Proving Equivalence of McGee- and Stalnaker$^+$-Validity

What is now to be shown is this: *A McGee sentence is McGee-valid just in case its translation is Stalnaker$^+$-valid*. The proof begins with a proof of the left-to-right version of the statement, followed by a proof of the inverse.

3.1 Left to Right

If a McGee sentence is McGee-valid, then the translation of that sentence is Stalnaker$^+$-valid.

McGee shows that a McGee sentence is McGee-valid just in case it follows by the rules of truth-functional logic from his axioms A1 through A7. For any derivation of a McGee sentence from an empty premise set, using the derivation rules of truth-functional logic plus rules for instantiating McGee's axioms, there is (the left-to-right claim says) a corresponding derivation of the Stalnaker translation from an empty premise set, using the derivation rules of truth-functional logic plus rules for instantiating the translations of McGee's axioms. Since the derivation rules of truth-functional logic are obviously Stalnaker$^+$-valid, we only need to show that the translation of each of McGee's axioms is Stalnaker$^+$-valid.

McGee's axioms A1 through A7 are as follows:

(A1) $(A \Rightarrow (X \wedge Y)) \equiv ((A \Rightarrow X) \wedge (A \Rightarrow Y))$
(A2) $(A \Rightarrow (X \vee Y)) \equiv ((A \Rightarrow X) \vee (A \Rightarrow Y))$
(A3) $(A \Rightarrow Y) \vee ((A \Rightarrow \neg X) \equiv \neg(A \Rightarrow X))$
(A4) $(A \Rightarrow (B \Rightarrow X)) \equiv ((A \wedge B) \Rightarrow X)$
(A5) $(A \wedge (A \Rightarrow B)) \equiv (A \wedge B)$
(A6) $((A \Rightarrow B) \wedge (B \Rightarrow A) \wedge (A \Rightarrow C)) \supset (B \Rightarrow C)$
(A7) $A \Rightarrow A$

In the Stalnaker translations, all $X \Rightarrow Y$ schemas are replaced by $\Box X > Y$:

(T1) $(\Box A > (X \wedge Y)) \equiv ((\Box A > X) \wedge (\Box A > Y))$
(T2) $(\Box A > (X \vee Y)) \equiv ((\Box A > X) \vee (\Box A > Y))$
(T3) $(\Box A > Y) \vee ((\Box A > \neg X) \equiv \neg(\Box A > X))$
(T4) $(\Box A > (\Box B > X)) \equiv (\Box(A \wedge B) > X)$
(T5) $(A \wedge (\Box A > B)) \equiv (A \wedge B)$
(T6) $((\Box A > B) \wedge (\Box B > A) \wedge (\Box A > C)) \supset (\Box B > C)$
(T7) $\Box A > A$

It is easily verified that T1, T2, T3 and T7 are Stalnaker-valid and therefore Stalnaker$^+$-valid.

T4 is Stalnaker$^+$-valid. Given reflexivity and transitivity of R and agreement under I, $V_S(\Box A)$ is forward-closed under R and closed under fusion. Consequently, by iteration $f(\Box B, f(\Box A, w)) = f(\Box A \wedge \Box B, w) = f(\Box(A \wedge B), w)$. So for any w, $f(\Box B, f(\Box A, w))$ is in

$V_S(X)$ just in case $f(\Box(A \wedge B), w)$ is, or else $f(\Box B, f(\Box A, w))$ and $f(\Box(A \wedge B), w)$ are both undefined. One way or another, any w is in V_S-of-axiom-T4. So V_S-of-T4 is W.

For proving the Stalnaker$^+$-validity of T5 and T6, the following lemma is useful.

Mirror lemma: For Boolean sentence A, let T_A be the set of terminal worlds in $V_S(A)$. Then for any world w such that $f(\Box A, w)$ is defined: with regard to accessibility and selection of worlds in T_A, $x = f(\Box A, w)$ mirrors w, meaning that x has access to exactly the same worlds in T_A as w does, and $f(T_A, x) = f(T_A, w)$ (or both sides of the equation are undefined). Proof: By transitivity, w has access to every world in T_A that x does. Now let z be an arbitrary world in T_A such that wRz. Then xRz, because otherwise there would by fusion of x and z with respect to w be a v such that wRv, vRx and vRz; v would be in $V_S(\Box A)$ since the only terminal worlds it would have access to would be ones in $V_S(A)$; and by monotonicity v, not x, would be $f(\Box A, w)$. And since $V_S(\Box A)$ is forward-closed and closed under fusion (as noted in the proof of T4's Stalnaker$^+$-validity), by iteration $f(T_A, f(V_S(\Box A), w)) = f(V_S(\Box A) \cap T_A, w)$, and so $f(T_A, x) = f(T_A, w)$, since $T_A \subseteq V_S(\Box A)$.

With this lemma in hand, we now show that T5 and T6 are Stalnaker$^+$-valid.

T5 is Stalnaker$^+$-valid. The biconditional is true at any world not in $V_S(A)$. Now consider arbitrary world w in $V_S(A)$. From w a world in $V_S(\Box A)$ is accessible, since by agreement w's closest accessible terminal world is in $V_S(A)$ and so in $V_S(\Box A)$. So now consider $x = f(\Box A, w)$, knowing that it is defined and is in $V_S(A)$. If T is the set of terminal worlds, then by agreement $f(T, x)$, like $f(T, w)$, is in $V_S(A)$, and then by the mirror lemma $f(T, w) = f(T_A, w) = f(T_A, x) = f(T, x)$. From this it follows, again by agreement, that either both or neither of w and x are in $V_S(B)$, and in either case the biconditional is true. Since w is arbitrary, V_S-of-T5 is W.

T6 is Stalnaker$^+$-valid. Consider an arbitrary world w. If $f(\Box B, w)$ is undefined, then w is in V_S-of-T6 because it is in $V_S(\Box B > C)$. If $f(\Box B, w)$ is defined but $f(\Box A, w)$ is not, then there is no terminal world, and hence by agreement and transitivity no world x of any kind, in $V_S(A)$ such wRx, and so $f(\Box B, w)$ is not in $V_S(A)$, so that w is not in $V_S(\Box B > A)$, and so again w is in V_S-of-T6, because it is not in V_S-of-the-antecedent. So now suppose that $f(\Box A, w)$ and $f(\Box B, w)$ are both defined, and suppose under these conditions that w is in $V_S(\Box A > B)$ and $V_S(\Box B > A)$, which is the only way w could fail to be in V_S-of-T6. Then $f(\Box A, w)$ and $f(\Box B, w)$ are each in both $V_S(A)$ and $V_S(B)$. If T is the set of terminal worlds, then let $z_1 = f(T, f(\Box A, w))$ and $z_1 = f(T, f(\Box B, w))$. By agreement, z_1 and z_2 are each in both $V_S(A)$ and $V_S(B)$. By the mirror lemma, $z_1 = f(T \cap V_S(A), f(\Box A, w)) = f(T \cap V_S(A), w)$, so it follows that $z_1 = f(T \cap V_S(A) \cap V_S(B), w)$. And similarly, $z_2 = f(T \cap V_S(A) \cap V_S(B), w)$, so $z_1 = z_2$. Whether or not this world is in $V_S(C)$, by agreement $f(\Box A, w)$ and $f(\Box B, w)$ are either both in $V_S(C)$ or else neither is, so w is in either both or neither of $V_S(\Box A > C)$ and $V_S(\Box B > C)$, and in either case w is in V_S-of-T6. Since w is arbitrary, V_S-of-T6 is W.

3.2 The Inverse

Now for the inverse of the statement just proved. *If a McGee sentence is not McGee-valid, then the translation of that sentence is not Stalnaker$^+$-valid.*

In McGee's canonical models (p. 515), W is finite. So for any sentence X, if there is a McGee model for which W is in $V_M(X)$, then there is a finite McGee model of the same description, i.e., with finite W in $V_M(X)$; and similarly for cases where W is not in $V_M(X)$.

Now, using prime superscripts to indicate the transition from McGee's syntax and semantics to their Stalnaker$^+$-counterparts, we represent the translations of X and Y as X' and Y', and for a finite McGee model $\langle W, R, f, w_0, I\rangle$, we construct a corresponding Stalnaker$^+$ model $\langle W', R', f', I'\rangle$.

- W' is the power set of W minus the empty set, $P(W) - \{\varnothing\}$, so that each 'world' in W' is a nonempty set of worlds that are in W.
- R' is the superset relation, which is reflexive.
- To construct f', we make use of the fact that f gives rise to a well-ordering of members of W, with $f(W, w_0) = w_0$ coming first, followed by $f(W - \{w_0\}, w_0)$, and so on until W is exhausted. This allows every member of W' to be treated as a well-ordered set. For any subset S of W' and member w of W', $f'(S, w)$ picks out that member x of S which is accessible from w and matches w more closely than any other member, $y \neq x$, of S that is also accessible from w, in this sense: the first member of w that is in x but not y or vice versa is in x. The binary relation between x and y of greater comparative similarity to w is asymmetric, transitive, and total. From this it follows that for models with finite W', $f'(S, w)$ is defined for any S containing a world accessible from w (there will be exactly one world in S that wins all pairwise similarity contests). Also, clearly if w itself is in S, then $f(S, w) = w$. Finally, when S_2 is a subset of S_1 that contains $f'(S_1, w)$, clearly $f'(S_2, w) = f'(S_1, w)$.
- I' is the function that maps each atomic sentence X' $(=X)$ to the set of members of W' whose first members are in $I(X)$. If, given I', V_S' is defined recursively as described earlier for Stalnaker models, then for any Boolean sentence X it follows that $V_S'(X)$ is the set of members of W' whose first members are in $V_S(X)$.

$\langle W', R', f', I'\rangle$ is a Stalnaker$^+$ model:

- R', the superset relation, is transitive; it is terminal, with singleton sets being the terminal members of W'; and for any w with subsets x and y, all members of W', the set $x \cup y$ is the unique fusion of x and y with respect to w. None of these properties are affected by the orderedness due to f.
- f' satisfies iteration. For any member w and any subsets S_1 and S_2 of W', where if any x is in S_1, then so are all its nonempty subsets, and if any x and y are in S_1, then so is $x \cup y$: $f'(S_2, f'(S_1, w)) = f'(S_1 \cap S_2, w)$, unless both sides of the equation are undefined. Proof: Suppose $x_1 = f'(S_1 \cap S_2, w)$ is defined; then x_1 is in both S_1 and S_2 and $x_1 \subseteq w$. It follows that $y = f'(S_1, w)$ must be defined (since w has an accessible S_1-world); and y must have access to x_1, meaning that $x_1 \subseteq y$, since otherwise $x_1 \cup y$, which must be in S_1 since both x_1 and y are, would be $f'(S_1, w)$, on account of containing every member of w that is in y; so that $f'(S_2, y) = f'(S_2, f'(S_1, w))$ is defined. Now suppose, conversely, that $x_2 = f'(S_2, f'(S_1, w))$ is defined; then since x_2 must be in S_2, and since x_2 is a subset of $y = f'(S_1, w)$, which is in S_1, which contains all

the nonempty subsets of any member, x_2 is in S_1, and $x_2 \subseteq w$, which means $S_1 \cap S_2$ contains a world accessible to w, so $f'(S_1 \cap S_2, w)$ is defined. Finally, suppose that $x_1 = f'(S_1 \cap S_2, w)$ and $x_2 = f'(S_2, f'(S_1, w))$ are distinct, with x_1 and x_2 each necessarily in both S_1 and S_2; then because $f'(S_1 \cap S_2, w) = x_1$, the first member of w that is in exactly one of x_1 and x_2 is in x_1, but because $f'(S_2, f'(S_1, w)) = x_2$, the first member of $f'(S_1, w)$ that is in exactly in exactly one of x_1 and x_2 is in x_2, which is impossible, since $f'(S_1, w)$ must contain all the members of w that are in either x_1 or x_2.

- I' satisfies agreement, because by the definition of f', for atomic X' a member w of W' is in $I'(X')$ iff the singleton set containing the first member of w is in $I'(X')$. That singleton set is $f'(T', w)$, where T' is the set of singleton subsets of W.

We now show that for any finite McGee model, McGee sentence X, and nonempty subset S of W, S is in $V_M(X)$ just in case for the corresponding constructed Stalnaker$^+$ model, S (considered as a member of W') is in $V_S'(X')$. We proceed by induction on the complexity of sentences.

The claim is true for any atomic sentence X, since S is in $V_M(X)$ iff $f(S, w_0)$ is in $I(X)$, iff S (whose first member under f is $f(S, w_0)$) is in $I'(X')$, iff S is in $V_S'(X')$.

Now suppose the claim is true for X and Y.

- S is in $V_M(\neg X)$ iff it is not in $V_M(X)$, iff it is not in $V_S'(X')$, iff it is in $V_S'(\neg X') = V_S'((\neg X)')$.
- S is in $V_M(X \wedge Y)$ iff it is in both $V_M(X)$ and $V_M(Y)$, iff it is both in $V_S'(X')$ and $V_S'(Y')$, iff it is in $V_S'(X' \wedge Y') = V_S'((X \wedge Y)')$.
- S is in $V_M(X \Rightarrow Y)$ iff $S \cap V_S(X)$ is in $V_M(Y)$, iff $S \cap V_S(X)$ is (a) in $V_M(Y)$ or (b) empty. Case (a): $S \cap V_S(X)$ is in $V_M(Y)$ iff $S \cap V_S(X)$ is in $V_S'(Y')$, which, since $f'(\Box X', S) = S \cap V_S(X)$, is true iff $f'(\Box X', S)$ is in $V_S'(Y')$. Case (b): $S \cap V_S(X)$ is empty iff no member of W is in both S and $V_S(X)$, iff no subset of S has its first member in $V_S(X)$, iff no member of member of W' accessible from S is in $V_S'(X)$, iff S is in $V_S'(\Box\neg X)$, iff $f'(\Box X, S)$ is undefined. So: S is in $V_M(X \Rightarrow Y)$ iff S is in $V_S'(\Box X' > Y') = V_S'((X \Rightarrow Y)')$.

So then for a McGee sentence X, if there is a McGee model for which W fails to be in $V_M(X)$, then there is a finite such model and a correspondingly constructed Stalnaker$^+$ model for which W, considered as a member of W', fails to be in $V_S'(X')$, so that $V_S'(X') \neq W'$. That is to say, if X is not McGee-valid, then X' is not Stalnaker$^+$-valid.

4 Interpretation and Concluding Remarks

We now turn to the interpretation of Stalnaker$^+$ models. Formally, each world in a Stalnaker$^+$ model can be thought of as a McGee model with its internal structure suppressed—a universe shrunk to a point, as it were. Section 3.2 explains how this conception of Stalnaker$^+$ worlds gives rise to the R, f, and I in Stalnaker$^+$ models. Terminality, fusion, and so on are not products of a brute trial-and-error search for properties that will support McGee's axioms. Rather, they reflect the conception of Stalnaker$^+$ worlds as encapsulated McGee models, intrinsically related to one another in certain ways.[2]

Substantively, the interpretation of Stalnaker$^+$ models is epistemic or doxastic. Instead of calling each index a possible world and speaking of sentences being true at, or in, worlds, let each index represent a possible *state of mind* of a hypothetical agent, and let sentences be *accepted* in states of mind. For any sentence X, in every state of mind either X or $\neg X$ is accepted. These are not ordinary states of mind, then, but highly dogmatic ones. Within each state, the dogmatism manifests as what Adams (1998, p. 134) calls an *order of magnitude probability ordering* of the McGee worlds: one world is far more likely than all the others, one world is by far the most-likely alternative to the first world, and so on.

However, if we want to use this semantics to model everyday reasoning, we can follow a plan like the one recommended by Stalnaker and Thomason (1970, p. 27n) and Stalnaker (1981, pp. 89–90) for validating Conditional Excluded Middle. The idea is to allow ties between worlds when it comes to order-of-magnitude probability and then to use a supervaluation approach (see van Fraassen 1966) to assess sentences for acceptance: a sentence is accepted in a state where ties between worlds occur just in case it would be accepted in every state where all ties were broken. Then a sentence may be accepted, rejected (its negation accepted), or neither, by an agent who considers some worlds plausible, others unlikely, yet others highly unlikely, etc. This means of getting three semantic values more or less for the price of two leaves the set of valid sentences unchanged. Because the logic is unaffected, the semantics of dogmatic states can be used to investigate normal, non-dogmatic reasoning.

Even the dogmatic states, however, are not utterly dogmatic, because an agent may accept X but be able to move, by *updating*, to a state in which $\neg X$ is accepted instead; this reachability of one state from another via updating is what the accessibility relation represents. If what is being updated on is information A (which we suppose to be factual, which in this context means Boolean), then the selection function picks out that accessible state which, of all the ones where A can count as having been updated on, represents the minimal alteration from the starting state. Conditional sentences invoke this updating operation: to say "If A…" is to contemplate a minimal shift to a state of mind in which A has been updated on.

Updating may be conceptualized in various ways, but here the paradigm is classical conditioning, which takes the information being updated on as certain. Therefore, updating on A mean moving not to the nearest state where A is accepted but, rather, to the nearest state where A is *irrevocably* accepted, which not coincidentally is the nearest state where $\Box A$ is accepted. Thus the state reached by updating is the nearest $\Box A$-state. If the nearest A-state is different, it would be reached by updating on something other than A. That is, updating on B might change the agent's mind

[2] There is one complication, which here can only be acknowledged, not dealt with: the construction in Sect. 3.2 assumes the set of McGee worlds to be finite. When it is infinite, the construction of the selection function for Stalnaker$^+$ models breaks down.

about A in such a way that further updating could change the agent's mind again—not about B, but about A.

All of the properties of the accessibility relation and the selection function have epistemic/doxastic interpretations. So, for example, the accessibility relation is terminal because at the end of every sequence of updatings there is a saturation point, where everything that is accepted is accepted irrevocably, and no further updating is possible, unless to fall into contradiction. Or take the fusion property: it holds because if updating on A would take the agent to state 1 while updating instead on B would take the agent to state 2, then there is a state arrived at by updating on $A \vee B$, from which either state 1 or state 2 could still be reached, via a second updating, on A or on B, respectively.

We conclude with some observations, which here are offered in abbreviated, provisional form, because it would be hard usefully to say more without saying a great deal more.

One observation concerns McGee-type failures of modus ponens. As is shown by the above proof that axiom T5 is Stalnaker^{+}-valid, the inference form $A, \Box A > B \,/\, B$, though not self-evidently valid, is valid due to the additional constraints on the accessibility relation and the selection function. But for non-Boolean X, it is, in general, perfectly reasonable for an agent to accept A and also $\Box A > X$, but not $\Box A$, and therefore be free not to accept X. On the other hand, it would be unreasonable for an agent to be *certain* of A, i.e., to accept $\Box A$, and also $\Box A > X$, but not X. In short, it is essential in McGee-style counterexamples to modus ponens that the unconditional premise be only revocably accepted. The Stalnaker translation of McGee conditionals makes this explicit.

A second observation concerns left-nesting, i.e., the construction of conditionals with conditionals in the antecedent. The proposed translation of McGee conditionals is premised on the assumption that antecedents are Boolean, but nothing stands in the way of our ignoring that assumption and translating, say, $(A \Rightarrow B) \Rightarrow C$ as $\Box(\Box A > B) > C$. Caution is in order, because the system is now being extended beyond what it was expressly designed for. But it is interesting that as things stand, what results is *not* McGee's earlier (1985) system, which permits left-nesting.[3]

Specifically, the Import–Export law, which McGee's earlier system is designed to support even with unrestricted left-nesting, fails. It is easy, for instance, to construct a three-state Stalnaker^{+} model with a state in which $(p \vee (p \Rightarrow q)) \Rightarrow (p \Rightarrow q)$ is accepted but $((p \vee (p \Rightarrow q)) \wedge p) \Rightarrow q$ is not—i.e., $\Box(p \vee (\Box p > q)) > (\Box p > q)$ is accepted but $\Box((p \vee (\Box p > q)) \wedge p) > q$ is not. The model contains terminal state 1, where p is rejected, and terminal state 2, where p is accepted and q is rejected. The state of evaluation is state 3, from which state 1 is closer than also-accessible state 2.[4]

The reason this is noteworthy is that, given the epistemic/doxastic interpretation of the semantics, the failure of Import–Export seems quite in order. Notice that (p

[3] Incidentally but relatedly: the one-fell-swoop replacement of every $\Rightarrow$ with a combination of $\Box$ and $>$ proposed here is quite different from McGee's recursive McGee-to-Stalnaker translation scheme (1985, pp. 469–470).

[4] Cantwell (2016) first drew my attention to the failure of Import–Export for left-nested constructions.

$\vee (p \Rightarrow q)) \wedge p$ is Boolean-equivalent to p, and that substitution of the latter for the former reduces $((p \vee (p \Rightarrow q)) \wedge p) \Rightarrow q$ to $p \Rightarrow q$. Surely someone who accepts $p \Rightarrow q$ on the assumption that either p or else $p \Rightarrow q$ is not thereby committed to accepting $p \Rightarrow q$ unconditionally.

To illustrate: imagine someone who lives in a locale where rain is rare and heavy rain is extremely rare. On a typical morning, this person doubts both $p =$ "It will rain" and $p \Rightarrow q =$ "If it rains, it will rain hard," and therefore doubts the disjunction of those two sentences. Called upon to suppose, nonetheless, that the disjunction is certain, the person must hypothetically either rule out sunny skies or else rule out light rain. The latter move is less of a mental adjustment, because light rain is rare anyway. Ruling out light rain makes $p \Rightarrow q$ certain. Therefore, the person will accept $(p \vee (p \Rightarrow q)) \Rightarrow (p \Rightarrow q)$. In natural language, this logical mouthful might come out as "On the assumption that either it will rain or else if it rains, it will rain hard, I say it will rain hard, if at all."[5]

At the same time, asked to suppose that either it will rain or else if it rains, it will rain hard *and also* that it will rain—invited, more simply put, to suppose that it will rain—our imagined person will *not* respond with "In that case, it will rain hard." Thus the Import–Export principle, which is rock-solid for Boolean antecedents (of indicative conditionals, that is—counterfactuals being another matter entirely), totters when left-nesting is introduced. Even if the present semantics ultimately turns out to be an imperfect tool for handling left-nesting, any improvement on it had best not make Import–Export unrestrictedly valid.[6]

Finally and very briefly: McGee has not only a logic but a scheme for assigning probabilities to sentences, including conditionals and arbitrary compounds thereof, using what amounts (as McGee notes, p. 517) to a quantitative version of the supervaluation approach outlined earlier. Not all Stalnaker$^+$ models can support an appropriately translated version of that scheme, but ones with the right mix of terminal and non-terminal worlds can. With those models, we can allocate the numerical probabilities in any arbitrary manner to the terminal worlds and then methodically redistribute the probabilities upstream, so to speak, to a (generally much larger) subset of the set of non-terminal states.

The particular redistribution method chosen by McGee is deterministic, meant to ensure that $\Pr(A \Rightarrow B) = \Pr(B \mid A)$ reliably holds, although $\Pr(A \Rightarrow X) = \Pr(X \mid A)$ can fail for non-Boolean X. The results are open to objections, such as those raised by Lance (1991). Be that as it may, what is interesting in the present context is that a modification of the probability distribution to update on factual information now takes place via what Lewis calls imaging (1976, p. 310). Thus the present semantics provides a way to construe McGee's $\Rightarrow$ connective as what Lewis calls a

[5] The scenario actually suggests a Stalnaker$^+$ model with three terminal states, corresponding to no rain, light rain, and heavy rain, respectively, and four non-terminal states. This seven-state model is less trivial than the three-state model described earlier.

[6] The non-equivalence of $(p \vee (p \Rightarrow q)) \Rightarrow (p \Rightarrow q)$ and $((p \vee (p \Rightarrow q)) \wedge p) \Rightarrow q$ also runs counter to the principle Mandelkern (2018b) calls Restricted Import–Export. On the other hand, the present semantics supports the principle Mandelkern calls Nothing Added, which he criticizes McGee's (1985) system for not validating.

probability-revision conditional (p. 312), except that updating on A means transferring each state's probability to the nearest $\Box A$-state, not the nearest A-state.

Acknowledgements I am grateful to Vann McGee, J. Karel Lambert, Hannes Leitgeb, John Cantwell, Wolfgang Spohn, Branden Fitelson, Jeffrey Barrett and especially Peter Woodruff for helpful exchanges, and to two *Erkenntnis* referees for detailed comments that led to significant revisions, especially in the last section.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford, CA: CSLI Publications.
Cantwell, J. (2016). The logic of the indicative conditional. (Manuscript).
Lance, M. (1991). Probabilistic dependence among conditionals. *Philosophical Review, 100,* 269–276.
Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review, 85,* 297–315.
Mandelkern, M. (2018a). The case of the missing 'if': Accessibility relations in Stalnaker's theory of conditionals. *Semantics and Pragmatics*. https://doi.org/10.3765/sp.11.8.
Mandelkern, M. (2018b). Import-export and 'and'. *Philosophy and Phenomenological Research*. https://doi.org/10.1111/phpr.12513.
McGee, V. (1985). A counterexample to modus ponens. *Journal of Philosophy, 82,* 462–471.
McGee, V. (1989). Conditional probabilities and compounds of conditionals. *Philosophical Review, 98,* 485–541.
Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in logical theory* (pp. 98–112). Oxford: Blackwell.
Stalnaker, R. C. (1981). A defense of conditional excluded middle. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs* (pp. 87–104). Dordrecht: D. Reidel.
Stalnaker, R. C., & Thomason, R. H. (1970). A semantic analysis of conditional logic. *Theoria, 36,* 23–42.
van Fraassen, B. C. (1966). Singular terms, truth-value gaps, and free logic. *Journal of Philosophy, 63,* 481–495.