Counterfactuals and Arbitrariness

Moritz Schulz

Penultimated draft. Please quote from the published version (*Mind* 2014).

Abstract

The pattern of credences we are inclined to assign to counterfactuals challenges standard accounts of counterfactuals. In response to this problem, the paper develops a semantics of counterfactuals in terms of the epsilon-operator. The proposed semantics stays close to the standard account: the epsilon-operator substitutes the universal quantifier present in standard semantics by arbitrarily binding the open world-variable. Various applications of the suggested semantics are explored including, in particular, an explanation of how the puzzling credences in counterfactuals come about.

1 Credences in Counterfactuals

According to what may be called the *standard view*, a counterfactual is true just in case the consequent is true at *all* relevant worlds at which the antecedent is true. For instance, the counterfactual 'If kangaroos had no tails, they would topple over' would be true just in case *all* relevant worlds at which kangaroos have no tails are worlds at which they topple over (cp. Lewis 1973, p. 1). More precisely:

> **Standard Semantics.** A counterfactual $\ulcorner \phi > \psi \urcorner$ is true at a world w iff ψ is true at *all* relevant ϕ-worlds with respect to w.

The standard account can be further substantiated by illuminating the notion of relevance at play, for example, by analyzing it in terms of similarity in certain respects (cp. e.g. Lewis 1979). It can also be modified along various dimensions.[1] But to fix ideas, I will stick to its present version.

As highlighted by Edgington (2008) and Moss (forthcoming), there is a puzzling phenomenon concerning the epistemic properties of counterfactuals. It seems that our subjective probabilities do not track the standard truth conditions. The structure of the problem is perhaps most obvious in the case of a fair lottery. Given that I did not purchase a ticket, what should I think about the following counterfactual?

(1) If I had bought a lottery ticket, I would have lost.

[1] For instance, Lewis (1973, p. 19ff.) rejects the so-called *limit assumption* which I will accept (on its technical construal). Cp. the discussion by Edgington (1995, p. 259) and Bennett (2003, §70), who follows up on a suggestion by Stalnaker (1978).

Intuitively, I should take it to be quite likely that if I had bought a lottery ticket, I would have lost. So, my credence in this counterfactual should be quite high. But it seems that I should not be certain that the counterfactual is true, for we consider it possible that if I had bought a lottery ticked, I would have won. Thus, the intuitive answer to the question 'Would I have lost if I had bought a lottery ticket?' seems to be 'Probably, but perhaps not'.

This intuitive verdict conflicts with the standard account. As Hawthorne (2005) points out, the standard account typically predicts that the counterfactual is false, for among the many relevant antecedent-worlds, there will be one world at which my ticket wins. Hence, the consequent is not true at all relevant antecedent-worlds, which makes the counterfactual false according to standard semantics. Moreover, the features of the actual world grounding the fact that most but not all relevant antecedent-worlds are consequent-worlds are easily accessible, for they comprise not much more than that we are dealing with a fair lottery. For this reason, I could be certain that the counterfactual is false if the standard account were correct. But intuitively I cannot be certain. On the contrary, it seems I should be fairly confident that the counterfactual is true.

These intuitions display a stable pattern across variations of the case. For instance, if we enlarge the lottery, we are inclined to assign higher and higher credences to the counterfactual, although the credences continue to fall short of 1. In the same way, if we shrink the lottery, we tend to assign lower and lower credences to the counterfactual, but our credences remain positive. The extreme case of a lottery with two tickets is probably best illustrated with a coin toss example. Intuitively, we should not exclude outright the counterfactual

(2) If the coin had been tossed, it would have landed heads,

although the standard view suggests that the counterfactual is evidently false, for at about half of the relevant worlds the coin lands tails. Rather, it seems appropriate to adopt an agnostic stance by holding that we do not (and perhaps cannot) know whether the coin would have come up heads if it had been tossed.

As the only exception, the standard view correctly predicts the conditions under which we should be certain about a counterfactual. Consider:

(3) If Obama had lost, McCain would have won.

We can plausibly assume that McCain wins at all relevant worlds at which Obama loses. Moreover, we know everything to realize this. Thus, the standard account can explain certainty about this counterfactual.

What is the reasoning behind our assignments of credences to counterfactuals? It seems that the subjective probabilities we are inclined to assign to counterfactuals reflect a certain pattern which is coordinated by how many relevant antecedent-worlds we take to be consequent-worlds. On the assumption that the proportion of consequent-worlds among the relevant antecedent-words is x, we are inclined to take the counterfactual to be x-likely. So, we are inclined to assign a credence of about $1/2$ to a coin-toss counterfactual because we take about half of the relevant antecedent-worlds to be consequent-worlds. Similarly, we take a lottery counterfactual to be likely because we assume that most relevant antecedent-worlds verify the consequent. Finally, on the assumption that all relevant antecedent-worlds are consequent-worlds, we can be certain about the counterfactual in accordance with the standard account.

This way of evaluating counterfactuals reflects a certain principle structurally similar to Lewis's (1980) Principal Principle. To a first approximation, we may put it this way:[2]

(4) $P(A > B|\mathrm{Ch}(B|A^*) = x)) = x$.

The conditional chances $\mathrm{Ch}(B|A^*)$ in terms of which the constraint is formulated should be construed as targetting the chance of the consequent B under the counterfactual assumption of A by operating on the set A^* of relevant A-worlds.[3] In the finite case, a uniform chance function can then be seen as measuring the proportion of consequent-worlds among the relevant antecedent-worlds. Just as for Lewis's Principal Principle, we have to restrict the principle to credence functions P which have not absorbed any inadmissible evidence. Then the principle says: on the assumption that the counterfactual chances of the consequent given the antecedent are x, we should take the counterfactual to be x-likely. In order to set a discussion of nested conditionals aside, we should take the principle to be concerned only with non-nested counterfactuals.

If this is how we tend to evaluate counterfactuals, then not only the standard account is in trouble: we face a general puzzle for the assignment of truth conditions to counterfactuals. The above suggests that (i) the subjective probability of a counterfactual should sometimes be high while our subjective probability that the standard truth conditions obtain should simultaneously be low (cp. e.g. the lottery example), whereas (ii) if we are certain that a counterfactual is true, we should always be certain that the standard truth conditions obtain, for in order to be certain about the counterfactual, we need to be certain that all relevant antecedent-worlds are consequent-worlds.[4] Now, feature (i) suggests that the truth conditions of counterfactuals must be weaker than the standard truth conditions, for otherwise a high credence in a counterfactual would demand a high credence that the corresponding standard truth conditions obtain by having them as a logical consequence. On the other hand, feature (ii) seems to require that the truth conditions of counterfactuals should be at least as strong as the standard truth conditions, for otherwise certainty about a counterfactual would not always allow for certainty that the standard truth conditions obtain. This, then, is the puzzle for counterfactuals deriving from our intuitive assignments of credences: they seem to imply and not to imply the standard truth conditions.

The aim of the present paper is to develop a semantics of counterfactuals according to which our credences in counterfactuals can be explained in the standard way as epistemic attitudes towards a proposition or content expressed

---

[2] A similar idea has been proposed by Skyrms (1980). Cp. also Moss (forthcoming) and Williams 2012. See Schulz (forthcoming) for a discussion of how to make the constraint more precise.

[3] Cp. Leitgeb 2012a for a related account of the relevant kind of chances in terms of Popper functions.

[4] It is well known that indicative conditionals give rise to a similar (but somewhat different) puzzle. See Edgington 1986, Edgington 1995, p. 278ff., and also Stalnaker 1975.

by the counterfactual as a whole. For lack of space, I have to omit discussion of non-standard explanations of the phenomenon under consideration. For instance, I will not discuss the suppositional view of counterfactuals defended by Edgington (2008) and Barnett (2010). Let me also mention that it might be possible to adapt the restrictor view of conditionals as suggested by Lewis (1975) and further developed by Kratzer (1979, 1986) to explain the problematic data. Although it is not hard to see that the relevant phenomenon cannot be explained in terms of domain restriction, one could consider explaining it in terms of a generalized form of domain modification. I have to leave discussion of this and similar possibilities for another occasion.

The paper is mainly divided in two parts. I will start by developing an alternative semantics for counterfactuals in terms of the epsilon-operator which is supposed to arbitrarily bind an open world-variable. The epsilon-operator will be a substitute for the universal quantifier present in standard semantics. In the second part, the account will be applied to various issues in the debate about counterfactuals, first and foremost to the explanation of our credences in counterfactuals. Before closing, the proposal will be compared to related theories of counterfactuals.

2 A Proposal

The structure of Stalnaker's (1968) semantics can be seen as providing the basis for truth conditions which are weaker than the standard truth conditions, for if all relevant antecedent-worlds are consequent-worlds, then a particular relevant antecedent-world is a consequent-world (but not *vice versa*). Yet Stalnaker's original background assumption that there is always a single most relevant antecedent-world is implausible. Standard semantics seems to be right in assuming that there are usually many relevant antecedent-worlds. In this section, I will present a semantics which shares its structure with Stalnaker's semantics but also preserves the presumption of the standard account that there are typically many relevant antecedent-worlds.

The basic idea will be to substitute in the standard account the epsilon-operator for the universal quantifier. Intuitively, the epsilon-operator *arbitrarily* selects a world out of the set of relevant antecedent-worlds. A counterfactual will then be true if the consequent is true at the arbitrarily selected antecedent-world. These truth conditions are weaker than the standard truth conditions: if all relevant antecedent-worlds are consequent-worlds, then an arbitrarily selected antecedent-world is a consequent-world, but an arbitrarily selected antecedent-world may be a consequent-world without all antecedent-worlds being consequent-worlds. One can already see how this provides a solution to the puzzle posed by the evaluation of counterfactuals. On the one hand, we can think it likely that an arbitrarily selected antecedent-world is a consequent-world without thinking it likely that all antecedent-worlds are consequent-words. This can happen, for instance, when we think that most but not all antecedent-worlds are consequent-worlds. On the other hand, we can only know that an arbitrarily selected antecedent-world is a consequent-world if we are in a position to know that all antecedent-worlds are consequent-worlds.

The element of arbitrariness present in such a semantics can be motivated by paralleling 'would'-expressions with 'will'-sentences. It is a common linguistic view that 'would' is derived from 'will' and can be regarded as a specific kind of past tense form of 'will'.[5] For this reason, it is a natural assumption that the semantics of 'would' should be similar to the semantics of 'will'. As 'will' is of intermediate strength between future necessity and future possibility, the modal character of 'would' can be expected to be intermediate between counterfactual necessity and counterfactual possibility (cp. Sect. 3.3). Now, the semantics for 'will' can be fixed—in some way or other—by binding it to the one of the possibly many futures which will eventually be actual. A similar approach is not possible in the case of 'would'. With respect to the many relevant counterfactual futures, there is no distinguished one which ultimately outlasts the others, for counterfactual futures do not play themselves out. One way to resolve this conflict is to bind the semantics of 'would' to an arbitrary counterfactual future.[6] Before we can develop this line of thought in more detail, we will need to take a closer look at the epsilon-operator.

2.1 The Epsilon-Operator

The epsilon-operator was originally introduced by Hilbert as part of his finitistic program.[7] Nowadays, the epsilon-operator is sometimes employed in linguistics to account for *E*-type pronouns and definite noun phrases (see e.g. Slater 1986 and von Heusinger 2004). Philosophically, the epsilon-operator is of interest, for it is closely related to the use of schematic or arbitrary names in prominent systems of natural deduction. For this reason, issues concerning the epsilon-operator resemble issues in the philosophical debate about arbitrary names and reference to arbitrary objects.

Let us start by saying something about the syntax of the epsilon-operator. The epsilon-operator is a term forming operator on formulas. If $\ulcorner\phi(x)\urcorner$ is a formula, then $\ulcorner\epsilon x\phi(x)\urcorner$ is a term (and not a formula such as $\ulcorner\forall x\phi(x)\urcorner$). As a result, the expression $\ulcorner\phi(\epsilon x\phi(x))\urcorner$ is a well-formed formula. Importantly, the variable x is not free in an epsilon-term like $\ulcorner\epsilon x\phi(x)\urcorner$. The epsilon-operator acts in this respect like a quantifier which binds x similar to a definite description operator.

Semantically, an epsilon-term like $\ulcorner\epsilon x\phi(x)\urcorner$ can be thought of as denoting an arbitrarily selected object satisfying $\ulcorner\phi(x)\urcorner$. But in contrast to definite descriptions, no uniqueness assumption about ϕ is made: an epsilon-term like $\ulcorner\epsilon x\phi(x)\urcorner$ is not taken to be empty if there are many ϕ-s. To give a semantic model for epsilon-terms like $\ulcorner\epsilon x\phi(x)\urcorner$, we will need a function which selects an object out of the set of ϕ-s. The epsilon-term can then be taken to denote

[5] More generally, the modal auxiliary in the consequent of counterfactuals always seems to take a certain past tense form. Cp. e.g. Dudman 1991 and Iatridou 2000.

[6] Another option might be to leave the world-variable in counterfactuals open. Many thanks to Jeff Russell for pointing this out to me. Such an idea could be inspired by the open-variable view of future contingents proposed in Belnap and Green 1994 and Belnap et al. 2001, Ch. 6.

[7] See Hilbert 1922, Hilbert 1923, and also Ackermann 1924. A general discussion of the role of the epsilon-operator in Hilbert's finitistic program can be found in Zach 2003.

the object so selected (Schröter 1956, 46). Formally, a selection function is a function s with the property that $s(A) \in A$ (provided that A is non-empty). Then, if A is the extension of the formula $\ulcorner \phi(x) \urcorner$, we can say that $\ulcorner \epsilon x \phi(x) \urcorner$ denotes $s(A)$.

Based on the present idea, Asser (1957) has developed a comprehensive model theory for the epsilon-operator. To begin with, we take a first-order language and supplement it with the epsilon-operator. We may then define a model $\mathcal{M}$ for this language to be a triple $\langle D, I, s \rangle$ with the following properties:

- D is any non-empty set (the *domain* of discourse).
- I is an *interpretation function* which assigns to any constant an object in D, to any n-ary predicate an n-ary relation over D, and to any n-ary function symbol a function from the n-ary Cartesian product of D into D.
- s is a function $\mathcal{P}(D) \to D$ such that $s(A) \in A$ if $A \neq \emptyset$.

The first two components of such a model are familiar. What is new is the selection function: its sole purpose is to account for epsilon-terms.

The next step would be to define what it means for a formula ϕ to hold in such a model $\mathcal{M}$ relative to a variable assignment β (formally, $\mathcal{M}, \beta \models \phi$). As it turns out, the fact that the epsilon-operator is a term forming operator on formulas complicates the definition considerably. But the basic idea is simple: add to the standard definition a clause which says that epsilon-terms are assigned their semantic value by the selection function. More formally, assume that $\mathcal{M}$ is a model and β a variable assignment. If v is a variable and a an object in the domain, then $\beta\frac{a}{v}$ is a variable assignment which differs from β only in assigning a to the variable v. For a set of sentences Γ and a sentence ϕ, the expression $\ulcorner \Gamma \models \phi \urcorner$ is defined in the standard way. Now, if R_β is the denotation function which assigns to each term a referent, then we will define this function in a way such that the following equation is satisfied (von Heusinger 2004, p. 313):

$$R_\beta(\ulcorner \epsilon v \phi(v) \urcorner) = s(\{a \in D : \mathcal{M}, \beta\frac{a}{v} \models \phi(v)\}).$$

Thus, epsilon-terms are assigned referents by applying the selection function s to the set of objects satisfying the formula which makes up the epsilon-term.

Let me briefly comment on epsilon-terms which contain an empty formula. If ϕ is not true of any objects in the domain, then the epsilon-term $\ulcorner \epsilon v \phi \urcorner$ is still assigned a semantic value, namely $s(\emptyset)$. Of course, the selected object cannot satisfy ϕ. To some extent, this may be felt to be counterintuitive: how can a descriptive term be denoting if nothing satisfies the description? However, on the present stipulation, the semantics remains classical which spares us a lot of work.[8] But let me point out that in the application to counterfactuals below, nothing will depend on this convention.

[8] Asser 1957 also contains a treatment of epsilon-terms with the aid of partial selection functions which are undefined for the empty set.

Asser (1957) axiomatized the resulting logic and proved it to be sound and complete. As it turns out, the epsilon-operator is governed by two basic axioms. The first axiom for the epsilon-operator has already been implicitly assumed:

Epsilon. $\exists v\phi(v) \supset \phi(\epsilon v\phi(v))$.[9]

In words: If there is an object satisfying ϕ, then the arbitrary object $\epsilon v\phi(v)$ satisfies ϕ. This is very intuitive: an arbitrary ϕ should always be a ϕ provided that there is a ϕ. This axiom is a theorem according to the present semantics, for the denotation of an epsilon-term is always an element of the extension of the embedded formula as long as this extension is non-empty.

The second basic axiom is this:

Extensionality. $\forall v(\phi(v) \equiv \psi(v)) \supset \epsilon v\phi(v) = \epsilon v\psi(v)$.

Informally: If two formulas are materially equivalent, the corresponding epsilon-terms denote the same object. This ensures that the semantic value of an epsilon-term depends only on the extension of the embedded formula. It is clear that this axiom holds in the present semantics, for the selection function takes only the extension of the relevant formula as an argument.

It is a common thought that a canonical way of proving a universally quantified formula is to show of an arbitrarily selected object that it satisfies this formula. In the light of this, it is a natural question whether a universally quantified sentence $\ulcorner\forall x\phi(x)\urcorner$ is implied by $\ulcorner\phi(\epsilon x(x = x))\urcorner$. The question is whether 'Everything is a ϕ' is implied by 'An arbitrary object is a ϕ'. When we formalize this latter sentence, we somehow need to express that the selection of objects is unrestricted. A good way of doing so seems to be to take a tautologous formula such as $\ulcorner x = x\urcorner$ which is logically guaranteed to be true of everything.

It is not hard to see, though, that the implication does not hold in the present semantics. This is because all that is required for an epsilon-formula such as $\ulcorner\phi(\epsilon x(x = x))\urcorner$ to be true in a model is that the selection function picks an object which is ϕ. But this is possible without all objects being ϕ-s. To put the point slightly differently: an epsilon-formula such as $\ulcorner\phi(\epsilon x(x = x))\urcorner$ is construed as a formula about a particular object, namely $\epsilon x(x = x)$, and this object may be ϕ without all objects being ϕ.

Nevertheless, there is a more qualified relation between epsilon-terms and the universal quantifier which can do justice to the convention of proving a universally quantified formula by proving—given certain side conditions—of an arbitrary object that it satisfies the formula:

Epsilon-Terms and Generality. Let Γ be a set of sentences which do not contain the epsilon-operator and let $\phi(x)$ be a formula which does not contain the epsilon-operator either. Then

$$\Gamma \models \phi(\epsilon x(x = x)) \text{ iff } \Gamma \models \forall x\phi(x).$$

[9] Assume that $\ulcorner\exists v\phi(v)\urcorner$ is a closed formula (a sentence). If we allow it to be open, a proviso is necessary to avoid an unwanted interplay of the variables in the epsilon-term and the quantifiers in the original formula. Cp. Asser 1957, p. 42.

Less formally: If it follows from a set of sentences that an arbitrary object is ϕ, then it also follows that everything is ϕ and *vice versa*. The side conditions should be no surprise: they exclude that any particular assumptions about $\epsilon x\phi(x)$ enter the stage.

A remark on how the present observation can be proven. Take a model $\mathcal{M}$ which verifies Γ. Since Γ does not contain the epsilon-operator, any model which differs from $\mathcal{M}$ only in the choice of the selection function will verify Γ as well. So, $\mathcal{M}$ verifies $\ulcorner\phi(\epsilon x(x = x))\urcorner$ no matter which object is assigned to the epsilon-term $\ulcorner\epsilon x(x = x)\urcorner$. Since ϕ does not contain the epsilon-operator either, the extension of ϕ is independent of the semantic values of epsilon-terms. Hence, $\mathcal{M}$ verifies ϕ no matter which object is assigned to its free variable showing $\ulcorner\mathcal{M} \models \forall x\phi(x)\urcorner$. So, if $\Gamma \models \phi(\epsilon x(x = x))$, then $\Gamma \models \forall x\phi(x)$. The other direction is an immediate consequence of universal instantiation.[10]

The relation between epsilon-terms and generality is reflected in the epistemic profile of sentences containing epsilon-terms. We can be justified to a high degree that an arbitrarily selected object has a certain property without being equally justified in thinking that everything has this property. If we know that most but not all objects have a certain property, we are justified to a high degree that an arbitrarily selected object has this property. However, we are not justified in thinking that all objects have this property. The possibility of having high credence in $\ulcorner\phi(\epsilon x(x = x))\urcorner$ without having high credence in $\ulcorner\forall x\phi(x)\urcorner$ reflects the fact that a sentence of the form $\ulcorner\phi(\epsilon x(x = x))\urcorner$ does not imply $\ulcorner\forall x\phi(x)\urcorner$. The situation changes when we consider knowledge. If we are in a position to know that an arbitrarily selected object has a certain property, we are in a position to know that everything has this property (provided we do not know anything particular about the object in question). Logically, this reflects the fact that if we can prove $\ulcorner\phi(\epsilon x(x = x))\urcorner$ (without making any specific assumptions about the epsilon-term), we can prove $\ulcorner\forall x\phi(x)\urcorner$.

The epistemic profile of sentences containing epsilon-terms is exactly of the kind we are looking for. There are logically stronger sentences, namely the corresponding universally quantified sentences, such that (a) we can be justified in assigning a high subjective probability to the epsilon-sentences without assigning a high subjective probability to these stronger sentences and (b) we can only know the epsilon-sentences to be true if we are in a position to know the corresponding stronger sentences to be true. Thus, epsilon-terms stand to the corresponding universally quantified sentences in the same epistemic relation as counterfactuals seem to stand to the standard truth conditions which are specified by a universal quantifier. For this reason, it seems promising to see whether we can give a semantics of counterfactuals by using the epsilon-operator.

[10]One may also wonder whether the same kind of equivalence holds between $\ulcorner\forall x(\phi(x) \supset \psi(x))\urcorner$ ('All ϕ-s are ψ-s') and $\ulcorner\psi(\epsilon x\phi(x))\urcorner$ ('An arbitrary ϕ is a ψ'). But when the universal quantification is vacuously true, the epsilon-sentence need not hold. However, once Γ proves $\ulcorner\exists x\phi(x)\urcorner$, the equivalence holds. This problem is avoided in the present theorem, for in classical logic, $\ulcorner\exists x(x = x)\urcorner$ is a logical theorem and therefore implied by any set of assumptions.

2.2 Semantics for Counterfactuals

Before we start, let me introduce some notions familiar from standard semantics (cp. Lewis 1973, Ch. 2.7). Let $\mathcal{W}$ be a set of possible worlds. We may introduce a *selection function* f which takes us from a world $w \in \mathcal{W}$ and a set of worlds $A \subseteq \mathcal{W}$ to a set of possible worlds $f_w(A) \subseteq \mathcal{W}$. The selection function is intended to model the notion of relevance. It should not be confused with the selection function s employed in the semantics for the epsilon-operator. Moreover, let $[\![]\!]$ be an *interpretation function* which assigns to a sentence ϕ the set of worlds at which ϕ is true, that is $[\![\phi]\!] = \{w \in W : w \models \phi\}$. The interpretation $[\![\phi]\!]$ of a sentence ϕ specifies the *truth conditions* of the sentence. According to this conception of truth conditions, two sentences have the same truth conditions if they are true at exactly the same possible worlds. Now, given a sentence ϕ, the set $f_w([\![\phi]\!])$ can be thought of as the set of relevant ϕ-worlds with respect to w.

In this terminology, the standard truth conditions can be specified like this:

Standard Formal Semantics. Let ϕ, ψ be sentences and $w \in W$. Then

$$(w \models \ulcorner \phi > \psi \urcorner) \text{ iff } (\forall w' \in f_w([\![\phi]\!]) : w' \models \psi).$$

Now, equipped with the epsilon-operator, we are able to talk about arbitrary worlds in our semantics. For instance, given a world w and a set of worlds A, $\epsilon w'(w' \in f_w(A))$ is an arbitrary relevant A-world (relative to w). With this possibility in mind, the proposal I would like to make can be stated as follows:

Epsilon-Based Truth Conditions. Let w be any world and $\ulcorner \phi > \psi \urcorner$ a counterfactual. Then $w \models \ulcorner \phi > \psi \urcorner$ iff

$$(\exists w' \in f_w([\![\phi]\!])) \supset (\epsilon\, w'(w' \in f_w([\![\phi]\!])) \models \psi).$$

In words: A counterfactual is true at a world w if either there is no relevant antecedent world or, if there is one, the consequent is true at the arbitrarily selected relevant antecedent-world (relative to w).

A few remarks on how these truth conditions are fixed. The truth conditions are specified by a material conditional, the antecedent of which expresses that there is a relevant antecedent-world. If there is none, the counterfactual comes out true. Thus, the present truth conditions share with the truth conditions of standard semantics the feature that counterfactuals with an impossible antecedent are vacuously true. This consequence can be avoided if one is prepared to allow for various non-trivial *impossible* worlds. In the case of an impossible antecedent, the epsilon-operator could then select an antecedent-world from a set of relevant impossible worlds. The main part of the specification of truth conditions is contained in the consequent of the material conditional. The selection function f takes the world w and the set $[\![\phi]\!]$ of worlds at which the antecedent is true as arguments and provides the set $f_w([\![\phi]\!])$ of relevant antecedent-worlds. Now, the epsilon-operator arbitrarily selects one of these worlds. The whole clause is then true if the consequent is true at this arbitrarily selected world.

The present kind of truth conditions is indeed weaker than the standard truth conditions. Suppose that the standard truth conditions are satisfied. If this is because there are no relevant antecedent-worlds, then the arbitrary truth conditions are satisfied as well. Otherwise the set $f_w([\![\phi]\!])$ will be non-empty and all worlds in it will verify the consequent of the counterfactual. But then an arbitrarily selected world out of this set has to verify the consequent. So, the standard truth conditions imply the epsilon-based truth conditions. The converse will generally not be the case, for an arbitrarily selected antecedent-world may verify the consequent without all antecedent-worlds doing so.

2.3 The Logic of Counterfactuals

So far, I have given an outline of an alternative semantics for counterfactuals. What remains to be done is to show how the logic of counterfactuals can be approached within the present framework. In standard semantics, the logic of counterfactuals can be adjusted by imposing constraints on the selection function f which provides the set of relevant antecedent-worlds. Since the present semantics is given partly in terms of this function, we can influence the logic in this way as well by constraining the sets from which the epsilon-operator selects its referents. In addition, one may introduce a coordination constraint which is directly concerned with the epsilon-operator.

Let us start with an obvious constraint on the function f:

Antecedent-Worlds. Let w be any world and A a set of worlds. Then

$$f_w(A) \subseteq A.$$

Simply put: Relevant antecedent-worlds are antecedent-worlds. This secures the logical truth of $\ulcorner A > A \urcorner$, for if relevant antecedent-worlds are always antecedent-worlds, then an arbitrarily selected relevant antecedent-world will be an antecedent-world. If $A = \emptyset$, i.e. the antecedent is impossible, the constraint implies that $f_w(A) = \emptyset$ as well, i.e. that there are no relevant antecedent-worlds. This makes counterfactuals with an impossible antecedent vacuously true.

The next constraint requires that the selection function selects some antecedent-worlds unless there are no antecedent-worlds.

Non-Emptiness. Let w be any world and A a non-empty set of worlds. Then

$$f_w(A) \neq \emptyset.$$

In words: if there are antecedent-worlds, then there are relevant antecedent-worlds. This constraint is important for our specification of truth conditions, for it guarantees that only counterfactuals with an impossible antecedent are vacuously true.

Two plausible constraints of increasing strength are known as *weak* and *strong centering* (cp. Lewis 1973, Sect. 1.7).

Weak Centering. Let w be any world in A. Then

$$w \in f_w(A).$$

Strong Centering. Let w be any world in A. Then

$$\{w\} = f_w(A).$$

The centering conditions are usually imposed to ensure the validity of *modus ponens*. On the present semantics, we would have to adopt strong centering. In this case, if the antecedent is true at the actual world, the selected world will be the actual world as it is the only possible choice and so the truth of the corresponding counterfactual will require the truth of its consequent.[11] Weak centering, on the other hand, would leave open the possibility that the selected antecedent-world is not the actual world and so modus ponens could fail. If one favours weak centering over strong centering as a description of the structure of relevance, one would need to impose an additional coordination constraint on the epsilon-operator (see below) or else accept failures of *modus ponens* for counterfactuals (which strikes me as the least attractive option).

If one compares the logic reached so far to Stalnaker's (1968) logic C2 or to Lewis's (1973) official logic VC (which is in other respects weaker than the present logic), then it turns out that there is precisely one interesting principle missing from the present logic. This principle states that counterfactually equivalent propositions can be substituted for each other in the antecedent of a counterfactual:

Substitution. $A > B, B > A, A > C \models B > C$.

This principle is quite complex and cannot easily be tested against intuitions. After an early attempt by Tichy (1978) to give a counterexample and a response by Stalnaker (1984, Ch. 7), this principle and its semantic justification have recently been put to closer scrutiny. Tooley (2002) discovered a certain type of counterexample to it the consequences of which have been further explored by Cross (2006).[12] The counterexample by Tooley requires, however, the possibility of backward causation. If one takes the counterexample to be genuine, one might simply stop here and leave the logic as it is (cp. Cross 2009).

Although I find the counterexample by Tooley *prima facie* very convincing, there are two considerations which speak in favour of having a backup plan. The first one is simply that it would be nice to prevent a semantics of counterfactuals from being hostage to fortune that backward causation is really possible. A related thought starts with the observation that there will in any case be large parts of metaphysical space which are free from backward causation. Unless the counterexample can be generalized to cases not involving backward causation, we would expect the substitution principle to hold in

[11] On a semantics like the present one which binds the truth-value of a counterfactual to what is true at a single world, the validity of *modus ponens* has the consequence that the actual truth of antecedent and consequent is sufficient for the truth of the corresponding counterfactual ($(A\&B) \supset (A > B)$). For a recent defence of this principle, see Walters 2009 and also Walters 2011 in response to Ahmed 2011.

[12] Ahmed (2011) also proposes a counterexample. But see Walters 2011 for a case against it. Further discussion which is particularly relevant for the epsilon-based semantics can be found in Bacon (ms).

such a restricted metaphysical environment. A good theory of counterfactuals should have the resources to bolster this expectation.

In standard semantics, the substitution principle is validated because the selection function satisfies the following constraint:

> **Equivalence for Relevance.** Let w be any world, A and B be sets of worlds. If $f_w(A) \subseteq B$ and $f_w(B) \subseteq A$, then
> $$f_w(A) = f_w(B).$$

In words: If the relevant A-worlds are B-worlds and the relevant B-worlds are A-worlds, then the relevant A-worlds coincide with the relevant B-worlds. Now, if the relevant A-worlds are the relevant B-worlds, then the truth of $\ulcorner A > C \urcorner$ implies the truth of $\ulcorner B > C \urcorner$ and so the substitution principle turns out to hold.

Actually, this equivalence condition is the consequence of a more general principle holding in standard semantics:

> **Generalized Equivalence.** Let w be any world, A and B be sets of worlds. If $f_w(A) \cap B \neq \emptyset$ and $f_w(B) \cap A \neq \emptyset$, then
> $$f_w(A) \cap f_w(B) = (f_w(A) \cap B) \cup (f_w(B) \cap A).$$

If the starting conditions of the original equivalence constraint are satisfied, then $f_w(A) \cap B = f_w(A)$ and $f_w(B) \cap A = f_w(B)$. Hence, the generalized principle would give us $f_w(A) \subseteq f_w(B)$ and $f_w(B) \subseteq f_w(A)$ showing that the original principle is indeed implied by the generalized one.[13]

To see why the generalized equivalence principle is a consequence of the standard account, recall that the closeness relation is assumed to be total in the sense that any two worlds are comparable with respect to their closeness to a given third world. The function f then takes a world w and a proposition A to the set of closest A-worlds with respect to w. Now, if some of the closest A-worlds are B-worlds and some of the closest B-worlds are A-worlds, these $(A\&B)$-worlds are all comparable to each other. If one of those worlds, w', were part of $f_w(A)$ but not of $f_w(B)$, say, then w' would either be closer to w than the $f_w(B)$-worlds or further away from them. In the first case, $f_w(B)$ would not consist of the closest B-worlds, for w is a B-world. In the second case, $f_w(A)$ would not consist of the closest A-worlds, for $f_w(B)$ contains A-worlds which are closer than w'. Thus, we see that the substitution principle effectively derives from the idea that the closeness relation is total combined with the general assumption that the truth of a counterfactual depends on what is true at the closest antecedent-worlds (cp. Cross 2006, 2008).

To clarify the relation between the substitution principle and the equivalence further, given the present semantics, suppose that the counterfactuals $\ulcorner A > B \urcorner$

[13] Hans Rott pointed out to me that the somewhat clumsy equivalence condition is provably equivalent to the conjunction of two more natural conditions known from rational choice theory and usually referred to as α and $\beta+$. I regret that I have not yet been able to explore these parallels further.

and $\ulcorner B > A \urcorner$ are true. This would mean that the arbitrarily selected relevant A-world is a B-world ($\epsilon w'(w' \in f_w(A)) \in B$) and the arbitrarily selected relevant B-world is an A-world ($\epsilon w'(w' \in f_w(B)) \in A$). Now, if we further assume that the counterfactual $\ulcorner A > C \urcorner$ is true, we get the consequence that $\ulcorner B > C \urcorner$ is true just in case our assumptions ensure the identity of the selected A-world with the selected B-world, for only then does the selected A-world being a C-world ensure that the selected B-world is a C-world as well. Now, if the equivalence constraint is in place, we get the required identity implication in the special case in which all relevant A-worlds are B-worlds and all relevant B-worlds are A-worlds. Then the relevant A-worlds will coincide with the relevant B-worlds by the equivalence constraint. Since these sets are identical, the selected relevant A-world will be the selected relevant B-world by the extensionality of epsilon-terms. However, if some of the relevant A-worlds are B-worlds but some are not, then the selected relevant A-world might be a B-world without all relevant A-worlds being B-worlds. In such a situation, the equivalence constraint will not be applicable and the relevant A-worlds will be distinct from the relevant B-worlds as some relevant A-worlds are not B-worlds. Still, some relevant B-worlds may be A-worlds and so the selected B-world may be an A-world without the selected worlds being identical. But then the substitution principle could fail.

The counterexamples to the substitution principle which would still occur even if the equivalence constraint on f is in place are somewhat peculiar. They can come about only in situations in which a counterfactual $\ulcorner A > B \urcorner$ is true without all relevant A-worlds being B-worlds. As explained below, the latter condition will prevent the original counterfactual $\ulcorner A > B \urcorner$ from being known (despite being true). Given that the additional counterexamples to the substitution principle would occur only in such specific circumstances, they may be tolerable despite lacking—as far as I can presently see—intuitive support (they cannot be justified in a way similar to Tooley's counterexample).

If one wishes to avoid these additional counterexamples to the substitution principle, one would need to introduce a new kind of *coordination constraint* directly concerned with epsilon-terms. One way of doing this is to constrain the choices open to the epsilon-operator in conditions pertaining on the validity of the substitution principle. In order to deal with possible counterexamples to the substitution principle, such a constraint should not be too strong and would ideally go hand in hand with the equivalence constraint for the relevance function f. Here is a suggestion:

> **Modest Equivalence for Choice.** Let w be any world, A and B be sets of worlds. If $\epsilon w'(w' \in f_w(A)) \in f_w(B)$ and $\epsilon w'(w' \in f_w(B)) \in f_w(A)$, then
> $$\epsilon w'(w' \in f_w(A)) = \epsilon w'(w' \in f_w(B)).$$

In words: If the arbitrary relevant A-world is a relevant B-world and the arbitrary relevant B-world is a relevant A-world, then the selected worlds are identical. If we combine this constraint with the generalized equivalence condition, we can show the substitution principle to hold. Suppose $\ulcorner A > B \urcorner$ and

$\ulcorner B > A \urcorner$ are true. Given our semantics, this requires that the arbitrarily selected relevant A-world is a B-world ($\epsilon w'(w' \in f_w(A)) \in B$) and the arbitrarily selected relevant B-world is an A-world ($\epsilon w'(w' \in f_w(B)) \in A$). By the generalized equivalence constraint, we find that the selected A-world is actually a relevant B-world and the selected B-world a relevant A-world by being in $f_w(A) \cap f_w(B)$. The modest equivalence condition on choice now gives us the identity of the selected worlds and so adding the premiss $\ulcorner A > C \urcorner$ would suffice for $\ulcorner B > C \urcorner$.

Adopting only something like the modest equivalence constraint on choice and feeding it in the generalized equivalence condition on relevance allows for a satisfying treatment of the substitution principle. If the substitution principle holds universally, the two constraints secure its validity. If, however, it has counterexamples like the one suggested by Tooley, this can be assumed to be the case because the equivalence condition on relevance fails: sometimes the relevant A-worlds are B-worlds and the relevant B-worlds are A-worlds, but the two sets of relevant worlds are none the less disjoint. In this situation, the antecedent for the modest equivalence condition on choice cannot be satisfied and so the substitution principle will fail. Accordingly, the likely truth of the substitution principle in most but possibly not all situations could be accounted for by the truth of the equivalence condition on relevance in most but not all situations.

Adding a coordination constraint like the present one to the picture may constitute a certain cost, for it substracts somewhat from the idea of having a choice operator which selects its referents on a completely random basis. Once a coordination constraint is in place, certain choices will no longer be independent of each other. For example, selecting a relevant A-world which happens to be a relevant B-world may require one to select this world also as the relevant B-world under certain circumstances. Clearly, the choices would not be determined, but the whole process would perhaps be better seen as the arbitrary selection of a whole structure representing an assignment of admissible choices rather than a number of (independent) individual choices. If one does not like the introduction of a coordination constraint, one could trade it in for a certain kind of additional counterexample to the substitution principle as described above. For the time being, I would like to leave open which line would be best to take.

3 Applications

Let me now put the present semantics to work. I will start by showing that it can be used to explain the puzzling credences we assign to counterfactuals. So, we may be able to give a fairly conservative solution of the puzzle by merely substituting the epsilon-operator for the universal quantifier in the standard semantics. Secondly, I draw some consequences for knowledge and assertion of counterfactuals which can be used to defend the account against the objection that conditional excluded middle turns out to be valid. Finally, I will discuss the duality thesis, partly because the semantics offers an interesting take on it and partly because the resulting failure of duality could be seen as a major disadvantage.

3.1 Credences in Counterfactuals

A major task for the present semantics is to explain the puzzling credences in counterfactuals reflected in the data presented earlier. The core idea is to explain our credences in counterfactuals by the specific epistemic profile of epsilon-terms: the more antecedent-worlds are consequent-worlds, the more likely should we take it to be true that an arbitrary antecedent-world is a consequent-world. I will only consider the most interesting case in which the antecedent of the counterfactual is false but not impossible.

Sentences containing epsilon-terms have a specific epistemic profile. Consider sentences of the form $\ulcorner G(\epsilon x F(x)) \urcorner$ saying that an arbitrary F is G. What are their subjective probabilities? It seems that our subjective probabilities concerning such sentences should be sensitive to what we take to be the distribution of Gs among the Fs provided we assume that there is at least one F. The more Fs we take to be Gs, the more likely should we take an arbitrary F to be a G. If we take few Fs to be Gs, we should think it quite unlikely that an arbitrary F is G. Similarly, if we take most Fs to be Gs, we should think it quite likely that an arbitrary F is G. In sum, how likely we should take an arbitrary F to be G seems to be coordinated with an estimate of the proportion of Gs among the Fs. Thus, epsilon-terms induce a certain structure on the subjective probabilities we should assign to sentences containing them: our subjective probabilities should be distributed uniformly over the different candidates which the epsilon-operator might select.

The relevant sentence in the specification of arbitrary truth conditions for counterfactuals is of the form $\ulcorner G(\epsilon x F(x)) \urcorner$. It reads: $\epsilon w'(w' \in f_w(\llbracket \phi \rrbracket)) \models \psi$, saying that an arbitrary relevant antecedent-world is a consequent-world. Following the above analysis of such epsilon-sentences, how likely we should take an arbitrary relevant antecedent-world to be a consequence-world is coordinated by an estimate of the proportion of consequent-worlds among the relevant antecedent-worlds. The more relevant antecedent-worlds we take to be consequent-worlds, the more likely should we take an arbitrary antecedent-world to be a consequent-world. For this reason, the present account conforms to the data concerning the subjective probabilities of counterfactuals.[14]

To see the advocated explanation at work, let us quickly go back to our original examples:

(5) If a given coin had been tossed, it would have landed heads,

(6) If I had bought a lottery ticket, I would have lost,

(7) If McCain had won, Obama would have lost.

[14] A discussion of possible triviality results for counterfactuals in the spirit of Lewis 1976, 1986b as pursued in Leitgeb 2012a and Williams 2012 lies beyond the scope of the present paper. A possible response to the problem is sketched in Schulz (ms). See also Bacon (ms), who works on a general defence against triviality based on the idea of random selection.

We found that we do not rule out the first counterfactual. Since some of the relevant toss-worlds are heads-worlds, we should not rule out that an arbitrary toss-world is a heads-world. That we take the second counterfactual to be probable is because we should have a high credence that an arbitrarily selected antecedent-world is a consequent-world given that most antecedent-worlds are consequent-worlds. Finally, the last counterfactual is such that all relevant antecedent-worlds are consequent-worlds. In this case, we can be certain that an arbitrary antecedent-world is a consequent-world.[15]

3.2 Knowability and Assertability

The present semantics for counterfactuals predicts a certain limitation of our epistemic access to counterfactuals. As we have seen, we can only know an arbitrary F to be G if all Fs are Gs. If only some Fs are Gs, we cannot know whether an arbitrarily selected F is G, for we cannot know which object has been selected. By the specification of truth conditions in terms of the epsilon-operator, it follows then that we can—special cases aside—only know that a counterfactual is true if all relevant antecedent-worlds are consequent-worlds.[16] We may therefore conjecture the following constraint to hold:

> **Unknown Counterfactuals.** If for some w' in $f_w([\![\phi]\!]) : w' \not\models \psi$, then the counterfactual $\ulcorner \phi > \psi \urcorner$ is in w not known to be true.

An important point is not captured by the present constraint. To see this, note that counterfactuals would satisfy the constraint even if they had the standard truth conditions. However, this would be so because the counterfactual would be false under the specified conditions and for this reason unknown. The situation is interestingly different with respect to the truth conditions in terms of the epsilon-operator. A counterfactual may be true at a world w because the arbitrary antecedent-world is a consequent-world despite the fact that not all relevant antecedent-worlds are consequent-worlds. In such a case, the counterfactual will be true but unknown.

Such a limitation of our epistemic access to counterfactuals squares well with the amount of counterfactual knowledge we are willing to ascribe to ourselves. A paradigm example of a counterfactual which, in typical circumstances, turns out to be undecided is a counterfactual like 'If the coin had been tossed, it would have landed heads'. We do not take ourselves to know that this counterfactual is true. More importantly, we do not take ourselves to know that this counterfactual is false either as one might expect on the standard account. Rather,

[15] I omit discussion of cases, involving e.g. biased coins, which can be taken to suggest that our credences are not always uniformly distributed over the relevant antecedent-worlds. I reckon that such cases call for a generalization of the present view which allows epsilon-operators being indexed with a possibly non-uniform probability distribution representing weighted choices. The present account would be the special case in which the indexed probability function is uniform.

[16] Exceptions are counterfactuals of the form 'If ϕ had been true, then the world would have been exactly like the arbitrarily selected relevant ϕ-world'. They can be known without all relevant antecedent-worlds verifying the consequent. Thanks to Timothy Williamson for pointing this out to me.

we are happy to say that we do not know *whether* the coin would have landed heads if it had been tossed. Similarly in the other cases we have discussed.

Although the present account implies a certain kind of skepticism concerning certain counterfactuals, it should be noted that there is no obstacle from the present point of view to the thesis that many counterfactuals can be known and are known to be true. In whichever situation the standard semantics predicts that a counterfactual is true because all relevant antecedent-worlds are consequent-worlds, the present account is compatible with the counterfactual being known. Since there are still many cases in which the standard truth conditions are satisfied, there are still many cases in which a counterfactual may be known. To some extent, the standard truth conditions turn out to approximate the knowability conditions of counterfactuals. This helps to explain the success of standard semantics (cp. Edgington 2008, p. 19).

As it stands, the principle concerning epistemic access to counterfactual facts does not have any modal force. It only states that certain counterfactuals are not known under certain circumstances. Does the reasoning behind this principle suggest anything stronger? A stronger claim would be that the counterfactuals under consideration cannot be known in the strong sense of it being metaphysically impossible that they are known. But such a stronger claim is not warranted. The present semantics does not imply that there are true counterfactuals which are strictly unknowable. The reason why this is so is simple: nothing in the present semantics suggests that if for a world w neither all worlds in $f_w(A)$ are ψ-worlds nor all worlds in $f_w(A)$ are $\neg\psi$-worlds, then there is no world w' such that either all worlds in $f_{w'}(A)$ are ψ-worlds or all worlds in $f_{w'}(A)$ are $\neg\psi$-worlds. But if there is such a world w', then it is compatible with the present semantics that the corresponding counterfactual is known in w' despite the fact that it is not known in w. An example might help to illustrate this. Suppose in the actual world a given lottery is fair. Then the semantics suggests that the counterfactual 'If I had bought a lottery ticket, I would have lost' is unknown. This is, of course, compatible with there being a possible world in which the lottery is rigged such that all tickets are determined to lose. In such a world, I may know the counterfactual to be true by knowing of the manipulation. So, the present account does not predict that the counterfactuals which are not known for semantic reasons are also strictly unknowable.

Let me also say something about assertions of counterfactuals. Of course, what to say will to a large extent depend on a general outlook on assertion. To illustrate the situation with a particular example, I will here follow Williamson (1996; 2000, Ch. 11) in assuming that the norm of assertion is knowledge: one should assert something only if one knows it. Given the constraint on counterfactual knowledge, we can derive a constraint on the assertability conditions of counterfactuals:

> **Assertability of Counterfactuals**. If for some w' in $f_w([\![\phi]\!]) : w' \not\models \psi$, then the counterfactual $\ulcorner \phi > \psi \urcorner$ is not assertable in w.

A counterfactual can only be known if all relevant antecedent-worlds are consequent-worlds, and a counterfactual should only be asserted if it is known to be true. Thus, the constraint follows from the knowability conditions of counterfactuals together with the general assumption that only what is known should be asserted.

Now, it seems to be a datum that in typical situations one should neither assert outright a counterfactual like 'If the coin had been tossed, it would have landed heads' nor 'If the coin had been tossed, it would have landed tails'. But on the present account, one of the two counterfactuals will be true. Initially, the fact that neither of the two counterfactuals is assertable may be levelled as an objection against the present theory, for it may be taken to suggest that neither of the two counterfactuals is true. By the principle above, however, the present datum can be explained in a straightforward way: the two counterfactuals should not be asserted because they are not known to be true. In addition, it can also be explained why there is the feeling that the non-assertability of these counterfactuals is somewhat more robust than in other cases in which a sentence happens to be non-assertable in a given situation. This can be explained by the fact that in typical situations our epistemic access to such counterfactuals is limited for principled reasons: if neither all antecedent-worlds are consequent-worlds nor all antecedent-worlds are non-consequent-worlds, there is nothing we can do to find out whether such a counterfactual is true.

This line of thought is directly relevant to another issue. On the present account, there is a certain kind of flexibility concerning the logic of counterfactuals. By adding or omitting constraints on the relevance function or directly on the coordination of choices, the logic of counterfactuals can be influenced. However, there is also a limit to this flexibility: certain inferences will be valid even if no constraints are imposed. So, there is a certain *base logic* for counterfactuals which results already from the general shape of the present semantics. There is one important consequence of this base logic which is controversial in the debate about counterfactuals: the validity of *conditional excluded middle.*

Conditional Excluded Middle. $(\phi > \psi) \vee (\phi > \neg\psi)$.

In words: if A had been the case, B would have been the case, or, if A had been the case, B would not have been the case. This principle is valid on the present semantics: if the antecedent is impossible, then both counterfactuals are vacuously true and if the antecedent is not impossible, then the selected ϕ-world will either be a ψ-world or a $\neg\psi$-world, in which case either $\ulcorner(\phi > \psi)\urcorner$ or $\ulcorner(\phi > \neg\psi)\urcorner$ will be true. In contrast, conditional excluded middle is invalid according to standard semantics, for it may be that neither all relevant ϕ-worlds are ψ-worlds nor all ϕ-worlds are $\neg\psi$-worlds.[17]

As the discussion above indicates, the disjuncts of an instance of conditional excluded middle can both be robustly unassertable. This helps explain—in a symmetric way—why whole instances of conditional excluded middle seem to sound fine (Lewis 1973, 80) while there can simultaneously be something wrong with both disjuncts. Our intuitions against the disjuncts can—on the present account—be assumed to derive from their unassertability, whereas our intuitions in favour of conditional excluded middle rightly owe themselves to its validity, which makes the disjunction always assertable.

[17]See Cross 2009 and Williams 2010 for a recent defence of conditional excluded middle.

3.3 Duality

A prominent hypothesis is that 'would'-counterfactuals and 'might'-counterfactuals are duals of each other (Lewis 1973, p. 21f.). The duality thesis can be given by the following schema (the arrow '$\diamond\!\!\rightarrow$' represents 'might'-counterfactuals):

Duality. $(\phi > \psi) \equiv \neg(\phi \diamond\!\!\rightarrow \neg\psi)$.

For instance, the following two counterfactuals would be equivalent:

(8) If Terry had scored, Chelsea would have won.

(9) It is not the case that if Terry had scored, Chelsea might not have won.

By exchanging the negation sign, the truth of a 'might'-counterfactual is supposed to imply the falsity of the corresponding 'would'-counterfactual:

(10) If the coin had been tossed, it might have landed tails

implies, by duality, the falsity of the counterfactual

(11) If the coin had been tossed, it would not have landed tails.

This example also shows that the duality thesis fails on our account. It is clear that (a) if the coin had been tossed, it might have landed tails and (b) if the coin had been tossed, it might have landed heads. But our semantics has it that either (a) if the coin had been tossed, it would have landed heads or (b) if the coin had been tossed, it would have landed tails, for the arbitrarily selected antecedent-world is either a heads-world or a tails-world. In any case, either the (a)-pair of a 'might'-counterfactual and a 'would'-counterfactual or the corresponding (b)-pair constitutes a counterexample to the duality thesis.

However, duality is not unproblematic. Consider the following sentences (cp. Stalnaker 1978, p. 100f.):

(12) If I had bought a lottery ticket, I might have won, but I am fairly sure that if I had bought a lottery ticket, I would not have won.

(13) If I had bought a lottery ticket, I might have won, but it is likely that if I had bought a lottery ticket, I would not have won.

These sentences are clearly assertable in appropriate circumstances. This would not be so if the duality thesis were true, for this thesis implies that in each sentence the embedded 'would'-counterfactual is logically equivalent to the negation of the initial, unembedded, 'might'-counterfactual. But instances of a schema like 'p, but it is likely that $\neg p$' are, for reasons akin to Moore's paradox, never assertable. This is evidence against duality. Nevertheless, it would be nice to explain why the duality thesis appears attractive.

It does not seem difficult to give a quick preliminary account of 'might'-counterfactuals in the present framework. Since the semantics is based on a function selecting the set of relevant antecedent-worlds, we can, for example, incorporate the standard approach to 'might'-counterfactuals: a 'might'-counterfactual of the form $\phi \diamond\!\!\rightarrow \psi$ is true iff ψ is true in at least one relevant ϕ-world. More formally:

Counterfactual Possibility. $w \models (\phi \diamond\!\!\rightarrow \psi)$ iff $\exists w' \in f_w(\llbracket\phi\rrbracket) : w' \models \psi$.[18]

Of course, the standard truth conditions are dual to the truth conditions for 'might'-counterfactuals so defined.

With this in mind, let us get back to the question of why the duality thesis appears attractive. Here we may note an interesting phenomenon. Although a 'might'-counterfactual and a corresponding 'would'-counterfactual with the negation signs filled in can both be true, the present account has it that they cannot both be known to be true. For, if the relevant 'might'-counterfactual is true, there will be among the relevant antecedent-worlds a world at which the consequent of the 'would'-counterfactual is false. By the knowability constraint, the 'would'-counterfactual can therefore not be known. Hence, the conjunction with the 'might'-counterfactual cannot be known and would, by the knowledge account of assertion, not be assertable. In effect, this explanation deems sentences like

(14) If Terry had scored, Chelsea would have won, but Chelsea might (still) have lost (even) if Terry had scored.

to be Moore-paradoxical: they are consistent, but the second conjunct undermines the possibility of knowing the first one. Thus, although duality fails on the semantic level, it has a true projection on the pragmatic level: the relevant pairs of 'might'-counterfactuals and 'would'-counterfactuals are not co-assertable.[19]

One application of the failure of duality concerns the recent debate about counterfactuals and objective chance.[20] On the assumption of non-trivial objective chances, many 'might'-counterfactuals seem true. By duality, many 'would'-counterfactuals would come out as false. For instance, there being a small objective chance that q conditional on p generally seems to support the claim that if it had been the case that p, it might have been the case that q. Given a slight objective chance that the plate flies off sideways, it seems true that

(15) If I had dropped the plate, it might not have been broken.

By duality, we would need to infer the falsity of

(16) If I had dropped the plate, it would have been broken.

In an indeterministic world, many cases have the underlying structure of this example, which would force many 'would'-counterfactuals to be false.[21] In this

[18]On this account, a 'might'-counterfactual is false if its antecedent is impossible, but this effect could be removed by imposing an appropriate proviso. 'Might'-counterfactuals with a true antecedents may also call for a modification if strong centering is adopted.

[19]DeRose (1999) and Stalnaker (1978) also give pragmatic explanations of the phenomenon. However, their explanations differ from the present one in that they take 'might'-counterfactuals to be epistemic or quasi-epistemic expressing something like the epistemic possibility of the corresponding 'would'-counterfactual. In contrast, 'might'-counterfactuals can be construed as non-epistemic or ontic on the present semantics.

[20]See Hájek (ms), Hawthorne 2005 and Williams 2008.

[21]It is not quite clear to what extent the assumption of indeterminism is essential to this argument. Even in a deterministic world, many (but perhaps less) 'might'-counterfactuals of

way, the duality thesis threatens many counterfactuals with falsity. On the present account, there is no need to go to that extreme. Since the duality thesis fails, the truth of many 'might'-counterfactuals does not imply the falsity of many 'would'-counterfactuals. Although the plate might have flown off sideways, we can still be highly confident that it would have been broken.[22]

4 Comparisons

Structurally, the present account is a kind of Stalnaker semantics. Whether or not a counterfactual is true depends on whether the consequent is true at a single world. *Prima facie*, one difference is that Stalnaker's original theory is more naturally described with a definite description operator: the truth conditions of counterfactuals are given by what is true at *the most relevant antecedent-world*. In contrast, it is an integral part of the present theory that there are usually many relevant antecedent-worlds for the epsilon-operator to select. But we could reinterpret Stalnaker's theory by identifying the most relevant antecedent-world with the arbitrarily selected world. Would such a theory be different from the present model? As far as the modal profile of the proposition expressed by counterfactuals goes, the answer is quite clearly 'No': both theories suggest that a counterfactual $\ulcorner A > B \urcorner$ is true at a set of worlds of the form $\{w : s_w(A) \models B\}$, where the function s selects the relevant antecedent-world. However, the present theory contains a further component which equips counterfactuals with a specific epistemic profile provided by the specification of their truth conditions in terms of the epsilon-operator. The explanation of why we should think about counterfactuals the way we do relies on the relevant antecedent-world being arbitrarily selected from a larger set of relevant antecedent-world. This feature about how the truth conditions are fixed cannot be directly read off the modal profile assigned to counterfactuals for reasons similar to why we might think differently about 'Hesperus is Phosphorus' and 'Hesperus is Hesperus' despite the fact that the two sentences have the same modal profile.

The epsilon-operator has been used in linguistics to account for definite noun phrases (von Heusinger 2004). Given the similarity between 'if'-clauses and definite noun phrases (Schlenker 2004), the present theory can to some extent be seen as an extension of von Heusinger's treatment of definite noun phrases to counterfactuals. A crucial difference would be that in the case of noun phrases, contextual salience fully determines the referents of epsilon-terms, whereas in our case it is important that the selection process is thought of to be random in order to account for the uncertainty about counterfactuals.

The general idea of accounting for the connection between conditionals and conditional probabilities in terms of random selection has two prominent predecessors. Van Fraassen (1976) obtained a limited tenability result by extend-

the envisaged kind still seem to be true. Cp. Hájek (ms) and the deterministic cases discussed in Edgington 2008.

[22]But maybe this solution is still not good enough if we think that we can even know that the plate would have been broken. Thanks to Daniel Dohrn for drawing my attention to this problem. The situation might parallel possible knowledge of the future. Cp. Hawthorne and Lasonen-Aarnio 2009 and the reply in Williamson 2009.

ing models in such a way that the new order of worlds represents the relevant antecedent-world as being randomly selected from the set of antecedent-worlds.[23] This result is achieved by moving to a model in which every world is an infinite series of worlds in the original model and then adjusting the order of these worlds in such a way that the closest antecedent-world can be seen as representing an arbitrarily selected antecedent-world. One advantage of the present model is that it is much simpler, more general, and allows conditionals to live their lives in the original domain of ordinary possible worlds. Another implementation of the same general idea is the random variable approach pursued by Jeffrey (1991) and further explored by Jeffrey and Stalnaker (1994) (see also Kaufmann 2005). The observation is that a conditional probability can be seen as the expected value of a random variable which assumes intermediate values matching the relevant conditional probability at worlds at which the antecedent is false.[24] Given that credences in classical propositions can be seen as the expected value of a random variable which only takes the values 1 and 0 depending on whether the proposition is true or false at the world in question, the random variable approach can be seen as an extension of the classical framework of propositions. However, conditionals would no longer be bivalent. They could have all values in the interval $[0, 1]$. The advantage of the present account is that it is more conservative by allowing conditionals to behave like ordinary propositions.

Very recently, Leitgeb (2012a,b) has developed a theory of counterfactuals in a framework of objective conditional chances. The motivation for his theory mainly consists in making counterfactuals scientifically more respectable by connecting them with objective chances, while at the same time describing them as expressing general but exception tolerant propositions. The proposal is roughly this: a counterfactual is true iff the conditional chance of its consequent given its antecedent equals 1/is close to 1/is high (where the alternatives describe different variants of the theory). Leitgeb's suggestion stays close to the standard account but weakens the truth conditions of counterfactuals slightly in a way similar to the 'near-miss'-proposal considered by Bennett (2003, p. 249ff.), according to which a counterfactual is true iff most relevant antecedent-worlds verify the consequent.

As it stands, a theory of this kind faces the same challenge as more standard theories of counterfactuals (cp. Edgington 2008). Our credences in counterfactuals do not seem to estimate whether the corresponding counterfactual chances are 1 (or high). To see this, suppose we know that the relevant chances are clearly lower than the proposed threshold but still fairly high. A possible example might be derived from a lottery with not too many tickets. In such a

[23]As van Fraassen was concerned with a different thesis applying to indicative conditionals, we would still need to see whether his result can be adapted to the evaluation of counterfactuals.

[24]The random variable approach targets, like van Fraassen, only Stalnaker's (1970) thesis. But it seems to have a natural extension to counterfactuals: where the evaluation of indicative conditionals might consist in estimating a random variable corresponding to our present subjective probabilities, counterfactuals might be evaluated by estimating a random variable corresponding to a certain kind of objective chances.

case, we should be certain that the counterfactual is false on the proposed truth conditions and invest no credence in it at all. Intuitively, however, we would be fairly confident that the counterfactual is true, for example, that we would not have won had we bought a lottery ticket.

In response, Leitgeb suggests that the troublesome credences may well not be credences that the proposed truth conditions obtain but rather—following Adams (1975, 1976) and Edgington (1995, 2008)—credences arrived at under the counterfactual supposition of the antecedent. Thus, there would be two ways of engaging epistemically with a counterfactual: (i) estimating how likely the truth conditions are satisfied and (ii) estimating how likely the consequent is satisfied under the counterfactual supposition of the antecedent. In conversation, the second kind of credences might be pragmatically conveyed despite the fact that the first kind will always be literally communicated. The challenge for such a dual account would be to come up with cases where our epistemic attitudes do indeed reflect what is predicted by (i) on the proposed truth conditions. As far as I can see, evidence that counterfactuals play a two-fold role in how they interact with epistemic modals and epistemic attitude verbs has not yet been found, but I take this to be an important question for future researchs.

5 Conclusion

The main question we have started with was: what explains that we evaluate counterfactuals the way we do? According to the present theory, reasoning about counterfactuals is akin to reasoning about arbitrary objects. We take a counterfactual to be likely to the extent that an arbitrary antecedent-world is a consequent-world. The semantics in terms of the epsilon-operator also shows that the way counterfactuals are evaluated is not necessarily in conflict with counterfactuals expressing propositions. It escapes the argument against truth conditions for counterfactuals by imposing certain limits on our epistemic access to counterfactual facts. In this sense, the theory involves a partial skepticism concerning counterfactuals: we are often not in position to know whether a given counterfactual is true or false. Nevertheless, the theory stays very close to standard semantics. Formally, it differs from standard semantics only in its employment of the epsilon-operator instead of the universal quantifier.

The proposed semantics yields a somewhat new picture of counterfactuals. Whereas counterfactuals are in standard semantics conceived of as expressing a certain kind of *restricted necessity*—truth of the consequent in *all* relevant antecedent-worlds—, the present theory describes counterfactuals as being of intermediate strength: weaker than counterfactual necessity but stronger than counterfactual possibility. In this way, counterfactuals resemble future contingents. In the case of the actual world, the truth-value of a future contingent is eventually settled by how the world happens to unfold. Something similar is not possible in the counterfactual case: we cannot wait and see how a counterfactual future evolves. This may be the source of why counterfactual involve an element of arbitrariness: which counterfactual future fixes the truth-value of a counterfactual is to some extent an arbitrary affair.

There remain various further issues. Perhaps the most important one concerns the metaphysical assumptions underlying the use of epsilon-terms in the present semantics. To address the issue properly, we would need to venture into the debate about arbitrary reference. In principle, any account of arbitrary reference may be adopted and applied to the present semantics of counterfactuals in terms of the epsilon-operator. There are at least two basic options. According to a fairly radical *realist* stance, there simply is a distinguished but arbitrary selection function which supplies epsilon-terms with referents.[25] In its application to counterfactuals, the realist stance would correspond to the assumption that there are some brute—in a certain sense arbitrary—counterfactual facts corresponding to the arbitrary selection of antecedent-worlds. This is the interpretation I have presupposed in the exposition of the theory so far and would be the one I favour if the metaphysical picture behind it does not turn out to be untenable. It should be noted, though, that a realist interpretation of the meta-semantic machinery is not the only option. For instance, it might be possible to underpin an *anti-realist* position by *supervaluating* over all possible selection functions which represent arbitrary choices or else to give a fictionalist story about epsilon-terms. On a supervaluationist interpretation, the present semantics gets close to Stalnaker's (1978) supervaluationist theory of conditionals, but may still have the advantage of being able to exploit the particular epistemic profile of epsilon-terms to explain the puzzling credences in counterfactuals (further research would be necessary to investigate to what extent the epistemological explanations given with a realist interpretation in mind carry over to an anti-realist account). Finally, let me point out that the present theory has been tailored to account for one specific problem the standard account has—the subjective probabilities of counterfactuals—and does not provide a remedy for various other issues such as reversed Sobel sequences (von Fintel 2001) and the integration of a theory of counterfactuals into a unified theory of conditionals.[26]

[25] A realist theory of arbitrary reference is defended in Breckenridge and Magidor 2012. An earlier presentation of this work at the Ockham Society in Oxford 2007 was an important source of inspiration for the present paper.

[26] **Acknowledgements.** Most of all, I would like to thank Dorothy Edgington and Benjamin Schnieder for their encouragement and support over the last five years, without which I would not have been able to pursue the present project. Special thanks are also due to the other members of the *phlox* research group: Nick Haverkamp, Miguel Hoeltje and Alexander Steinberg. Various friends and collegues have given me valuable comments on earlier drafts of the paper: Frank Arntzenius, Andrew Bacon, Daniel Dohrn, Ofra Magidor, Graham Priest, Tobias Rosefeldt, Sven Rosenkranz, Hans Rott, Lee Walters, and Timothy Williamson. I would also like to thank three anonymous referees for *Mind* and the editor of *Mind* for their helpful comments which led to significant improvements. Earlier drafts of the paper have been presented at the SIFA Graduate Conference 2007, the Oxford Graduate Conference 2007, the Ringvorlesung in Hamburg 2007, the Arche/CSMN Graduate Conference, the *phloxshop* on modality 2009, the PERSP metaphysics seminar 2011 and the G27 meeting 2013. Many thanks to all the participants for their stimulating comments. In writing this paper, I have benefited from partial funds by the Spanish Ministry of Science and Innovation for the project CONSOLIDER-INGENIO 2010 CSD2009-00056 on *Philosophy of Perspectival Thoughts and Facts (PERSP)* and for the project FFI2009-13436, I+D+i programme on *Semantic Content and Context Dependence*. I have also profited from the generous support of the Deutsche Forschungsgemeinschaft while working on an earlier draft of this paper.

References

Ackermann, Wilhelm 1924: 'Begründung des „tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit'. *Mathematische Annalen*, 93, pp. 1–36.

Adams, Ernest W. 1975: *The Logic of Conditionals*. Dordrecht: Reidel.

—— 1976: 'Prior Probabilities and Counterfactual Conditionals'. In Harper and Hooker 1976, pp. 1–21.

Ahmed, Arif 2011: 'Walters on Conjunction Conditionalization'. *Proceedings of the Aristotelian Society*, 111, pp. 115–22.

Asser, Günter 1957: 'Theorie der logischen Auswahlfunktionen'. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 3, pp. 30–68.

Bacon, Andrew ms: 'In Defence of a Naïve Conditional Epistemology'. Ms, in progress.

Barnett, David 2010: 'Zif Would Have Been If: A Suppositional View of Counterfactuals'. *Noûs*, 44, pp. 269–304.

Bäuerle, R., U. Egli, and A. von Stechow (eds) 1979: *Semantics from different points of view*, Berlin, Heidelberg, and New York: Springer.

Belnap, Nuel and Mitchell Green 1994: 'Indeterminism and the Thin Red Line'. *Philosophical Perspectives*, 8, pp. 365–88.

Belnap, N., M. Perloff, and M. Xu 2001: *Facing the Future. Agents and Choices in Our Indeterminist World*. Oxford: Oxford University Press.

Bennett, Jonathan 2003: *A Philosophical Guide to Conditionals*. Oxford: Oxford University Press.

Breckenridge, Wylie and Ofra Magidor 2012: 'Arbitrary Reference'. *Philosophical Studies*, 158, pp. 377–400.

Cross, Charles B. 2006: 'Conditional Logic and the Significance of Tooley's Example'. *Analysis*, 4, pp. 325–35.

—— 2008: 'Antecedent-Relative Comparative World Similarity'. *Journal of Philosophical Logic*, 37, pp. 101–120.

—— 2009: 'Conditional Excluded Middle'. *Erkenntnis*, 70, pp. 173–88.

DeRose, Keith 1999: 'Can It Be That It Would Have Been Even Though It Might Not Have Been?'. *Philosophical Perspectives*, 13, pp. 385–413.

Dudman, Victor H. 1991: 'Interpretations of 'If'-Sentences'. In Jackson 1991, pp. 202–32.

Edgington, Dorothy 1986: 'Do Conditionals Have Truth Conditions?'. *Critica*, 18, pp. 3–30.

—— 1995: 'On Conditionals'. *Mind*, 104, pp. 235–329.

—— 2008: 'Counterfactuals'. *Proceedings of the Aristotelian Society*, 108, pp. 1–21.

Eells, Ellery and Skyrms, Brian (eds) 1994: *Probability and Conditionals.* Cambridge: Cambridge University Press.

Farley, A. M., P. Farley, and K. E. McCollough (eds) 1986: *Papers from the Parasession on Pragmatics and Grammatical Theory.* Chicago: Chicago Linguistics Society.

Greenough, Patrick and Duncan Pritchard (eds) 2009: *Williamson on Knowledge.* Oxford: Oxford University Press.

Hájek, Alan ms: *Most Counterfactuals are False.* Ms, in progress.

Harper, William L. and Clifford A. Hooker (eds) 1976: *Foundations of Probability theory, Statistical Inference, and Statistical Theories of Science*, Vol. 1. Dordrecht: Reidel.

Harper, W. L., R. Stalnaker, and C. T. Pearce (eds) 1981: *Ifs.* Dordrecht: Reidel.

Hawthorne, John 2005: 'Chance and Counterfactuals'. *Philosophy and Phenomenological Research*, LXX, pp. 396–405.

Hawthorne, John and Maria Lasonen-Aarnio 2009: 'Knowledge and Objective Chance'. In Greenough and Pritchard 2009, pp. 92–108.

Hilbert, David 1922: 'Neubegründung der Mathematik: Erste Mitteilung'. *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, 1, pp. 157–77, rpnt (with notes by Bernays) in Hilbert 1935.

—— 1923: 'Die logischen Grundlagen der Mathematik'. *Mathematische Annalen*, 88, pp. 151–165, rpnt in Hilbert 1935.

—— 1935: *Gesammelte Abhandlungen*, Vol. 3. Berlin: Springer.

Iatridou, Sabine 2000: 'The Grammatical Ingredients of Counterfactuality'. *Linguistic Inquiry*, 31, pp. 231–70.

Jackson, Frank (ed.) 1991: *Conditionals.* Oxford: Clarendon Press.

Jeffrey, Richard 1991: 'Matter of Fact Conditionals'. *Proceedings of the Aristotelian Society Supplementary Volume*, 65, pp. 161–83.

Jeffrey, Richard and Robert Stalnaker 1994: 'Conditionals as Random Variables'. In Eells and Skyrms 1994, pp. 31–46.

Kaufmann, Stefan 2005: ‘Conditional Predictions. A Probabilistic Account’. *Linguistics and Philosophy*, 28, pp. 181–231.

Keenan, Edward L. (ed.) 1975: *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press.

Kenstowicz, Michael (ed.) 2001: *Ken Hale: A Life in Language*. Cambridge, MA: MIT Press.

Kratzer, Angelika 1979: ‘Conditional Necessity and Possibility’. In Bäuerle et al. 1979, pp. 117–47.

—— 1986: ‘Conditionals’. In Farley et al. 1986, pp. 115–135. Updated in Kratzer 2012.

—— 2012: *Modals and Conditionals*. Oxford: Oxford University Press.

Leitgeb, Hannes 2012a: ‘A Probabilistic Semantics for Counterfactuals. Part A’. *The Review of Symbolic Logic*, 5, pp. 26–84.

—— 2012b: ‘A Probabilistic Semantics for Counterfactuals. Part B’. *The Review of Symbolic Logic*, 5, pp. 85–121.

Lewis, David 1973: *Counterfactuals*. Oxford: Blackwell.

—— 1975: ‘Adverbs of Quantification’. In Keenan 1975, pp. 3–15.

—— 1976: ‘Probabilities of Conditionals and Conditional Probabilities’. *Philosophical Review*, 85, pp. 297–315.

—— 1979: ‘Counterfactuals and Time’s Arrow’. *Noûs*, 13, pp. 455–76.

—— 1980: ‘A Subjectivist’s Guide to Objective Chance’. In Harper et al. 1981, pp. 267–97.

—— 1986a: *Philosophical Papers*, Vol. 2. New York and Oxford: Oxford University Press.

—— 1986b: ‘Probabilities of Conditionals and Conditional Probabilities II’. *Philosophical Review*, 5, pp. 581–89.

Moss, Sarah forthcoming: ‘Subjunctive Credences and Semantic Humility’. Forthcoming in *Philosophy and Phenomenological Research*.

Schlenker, Philippe 2004: ‘Conditionals as Definite Descriptions (A Referential Analysis)’. *Research on Language and Computation*, 2, pp. 417–62.

Schröter, Karl 1956: ‘Theorie des bestimmten Artikels’. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 2, pp. 37–56.

Schulz, Moritz forthcoming: ‘Counterfactuals and Probability’. Forthcoming in *Logique et Analyse*.

—— ms: *Counterfactuals and Probability*. Monograph in preparation.

Skyrms, Brian 1980: 'The Prior Propensity Account of Subjunctive Conditionals'. *The University of Western Ontario Series in Philosophy of Science*, Vol. 15, pp. 259-265.

Slater, Hartley 1986: 'E-type Pronouns and ϵ-terms'. *Canandian Journal of Philsophy*, 16, pp. 27–38.

Stalnaker, Robert 1968: 'A Theory of Conditionals'. *Studies in Logical Theory*, 2, pp. 98–112.

—— 1970: 'Probability and Conditionals'. *Philosophy of Science*, 37, pp. 64–80.

—— 1975: 'Indicative Conditionals'. *Philosophia*, 5, pp. 269–86.

—— 1978: 'A Defense of Conditional Excluded Middle'. In Harper et al. 1981, pp. 87–104.

—— 1984: *Inquiry*. Cambridge, MA and London: MIT Press.

Stalnaker, Robert and Thomason, Richard 1968: 'Modality and Reference'. *Noûs*, 2, pp. 359–72.

Tichy, Pavel 1978: 'A New Theory of Subjunctive Conditionals'. *Synthese*, 37, pp. 433–57.

Tooley, Michael 2002: 'Backward Causation and the Stalnaker-Lewis Approach to Counterfactuals'. *Analysis*, 3, pp. 191–97.

van Fraassen, Bas 1976: 'Probabilities of Conditionals'. In Harper and Hooker 1976, pp. 261–301.

von Fintel, Kai 2001: 'Counterfactuals in a Dynamic Context'. In Kenstowicz 2001, pp. 123–52.

von Heusinger, Klaus 2004: 'Choice Functions and the Anaphoric Semantics of Definite NPs'. *Research on Language and Computation*, 2, pp. 309–29.

Walters, Lee 2009: 'Morgenbesser's Coin and Counterfactuals with true Components'. *Proceedings of the Aristotelian Society*, 109, pp. 365–79.

—— 2011: 'Reply to Ahmed'. *Proceedings of the Aristotelian Society*, 111, pp. 123–33.

Williams, J. Robert G. 2008: 'Chances, Counterfactuals and Similarity'. *Philosophy and Phenomenological Research*, 77, pp. 385–420.

—— 2010: 'Defending Conditional Excluded Middle'. *Noûs*, 44, pp. 650–68.

—— 2012: 'Counterfactual Trivaility: A Lewis-Impossibility Argument for Counterfactuals'. *Philosophy and Phenomenological Research*, 85, pp. 648–70.

Williamson, Timothy 1996: 'Knowing and Asserting'. *Philosophical Review*, 105, pp. 489–523.

—— 2000: *Knowledge and Its Limits*. Oxford: Oxford University Press.

—— 2009: 'Replies to Critics'. In Greenough and Pritchard 2009, pp. 279–384.

Zach, Richard 2003: 'The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's Program'. *Synthese*, 137, pp. 211–59.