THE REVIEW OF SYMBOLIC LOGIC
Volume 5, Number 1, March 2012

A PROBABILISTIC SEMANTICS FOR COUNTERFACTUALS. PART A

HANNES LEITGEB
Ludwig-Maximilians-University Munich

In memory of Horacio Arló-Costa

Abstract. This is part A of a paper in which we defend a semantics for counterfactuals which is probabilistic in the sense that the truth condition for counterfactuals refers to a probability measure. Because of its probabilistic nature, it allows a counterfactual 'if A then B' to be true even in the presence of relevant 'A and not B'-worlds, as long such exceptions are not too widely spread. The semantics is made precise and studied in different versions which are related to each other by representation theorems. Despite its probabilistic nature, we show that the semantics and the resulting system of logic may be regarded as a naturalistically vindicated variant of David Lewis' truth-conditional semantics and logic of counterfactuals. At the same time, the semantics overlaps in various ways with the non-truth-conditional suppositional theory for conditionals that derives from Ernest Adams' work. We argue that counterfactuals have two kinds of pragmatic meanings and come attached with two types of degrees of acceptability or belief, one being suppositional, the other one being truth based as determined by our probabilistic semantics; these degrees could not always coincide due to a new triviality result for counterfactuals, and they should not be identified in the light of their different interpretation and pragmatic purpose. However, for plain assertability the difference between them does not matter. Hence, if the suppositional theory of counterfactuals is formulated with sufficient care, our truth-conditional theory of counterfactuals is consistent with it. The results of our investigation are used to assess a claim considered by Hawthorne and Hájek, that is, the thesis that most ordinary counterfactuals are false.

§1. Introduction.

Speaking and reasoning in terms of counterfactuals are part and parcel of our manifest image of the world. However, our scientific image of the physical world seems to acknowledge just one kind of objective modality, that is, objective chance. In this article we attempt to bridge this divide between the manifest and the scientific image by introducing a semantics for counterfactuals that is based on chance. This will allow us to ground counterfactuals semantically in a scientifically respectable manner, and to make sure that many of the ordinary counterfactuals that we normally take to be true still come out as true or as approximately true according to the new semantics and our current scientific theories.

Actually, we could start this paper much in the same way as Vann McGee once began his article on "Conditional Probabilities and Compounds of Conditionals":

> Ernest Adams (1965, 1975) has advanced a probabilistic account of conditionals, according to which the probability of a simple English indicative conditional is the conditional probability of the consequent given the antecedent. The theory describes what English speakers assert and accept with unfailing accuracy, yet the theory has won only limited acceptance.

Received: August 10, 2010.

doi:10.1017/S1755020311000153

> A principal reason for this has been that the theory is so limited in its scope. (McGee, 1989, p. 485) [1]

Like McGee's article, the present paper also derives from Adams' (1965, 1975) seminal work on a probabilistic semantics for conditionals. Just as in McGee's article, this paper will state a probabilistic semantics for languages that are more inclusive than the ones considered by Adams, that is, for languages that allow for compound conditionals and Boolean combinations of conditionals.[2] However, whereas McGee's theory is best viewed as a semantics for indicative conditionals, in line with Adams' primary concerns, we suggest the semantics in the present paper to be a probabilistic semantics for counterfactuals or subjunctive conditionals written as usual as $A \mathrel{\Box\!\!\rightarrow} B$,[3] as in the paradigmatic

> If it were the case that A, then it would be the case that B

or

> If it had been the case that A, then it would have been the case that B

or

> If it were to be the case that A, then it would be the case that B.

So our focus is on conditionals which are formulated in the subjunctive mood; indeed, we will use the terms 'subjunctive' and 'counterfactual' synonymously, hence, when we speak of a counterfactual, we do not presuppose its antecedent to be false.[4] Moreover, unlike Adams' or McGee's theory, we will not just be interested in the acceptability or assertability conditions of conditionals, but also—quite conservatively—in their truth conditions. The truth condition for counterfactuals that we are going to defend,

> $A \mathrel{\Box\!\!\rightarrow} B$ is true in a world w iff the conditional chance of B given A at w is very high,

will be probabilistic in the sense that it refers to a probability measure, but the intended interpretation of any such probability measure will not be subjective epistemic and agent relative as in Adams' or McGee's theory, but rather conditional chances will be considered to be objective, nonepistemic, and world relative. Yet the semantics will touch on Adams' theory in various ways, as we shall see later.

We *do* regard the standard Lewisian semantics for counterfactuals (cf. Lewis, 1973a, 1973b) as excellent in many respects. Ultimately, one of our concerns will be to supply a

---

1 We should stress that since the publication of McGee's paper things have changed in so far as Adams' theory has won more than just "limited acceptance": mainly through Edgington (1995) and Bennett (2003), the Suppositional Theory of indicative conditionals has become the main competitor of the traditional "horseshoe analysis" of indicative conditionals if not in fact the front-runner.

2 We will even allow for conditionals with conditional antecedents, which goes beyond the scope of McGee's semantics. It is in fact possible to apply Adams' own semantics to languages with arbitrary Boolean combinations of conditionals (though without nested conditionals) without changing the original framework in any serious manner, as shown by Adams himself and by Schurz (1998).

3 We use Lewis' symbol $\mathrel{\Box\!\!\rightarrow}$ for the subjunctive 'if... then'. In contrast, we will use $\supset$ in order to express material conditionals and $\rightarrow$ to express indicative conditionals.

4 It is well-known, in fact, that there are explanatory uses of subjunctive conditionals in which the falsity or improbability of their antecedents is not implied in any sense; see Adams (1975, p. 111f) and Skyrms (1980b, p. 99).

(probabilistic) variant of it with scientifically respectable and accessible foundations, and also to make it slightly easier for some counterfactuals to be true in the new semantics than in Lewis' original semantics. In very rough terms: while we do not count $A \boxright B$ as true if and only if all closest A-worlds are B-worlds, but rather if and only if the conditional chance of B given A is close to 1, we will derive the latter to be equivalent to: in all closest worlds among those in which A has a chance to be true the conditional probability of B given A is close to 1.[5] This will have the consequence that there will be counterfactuals $A \boxright B$ which are true even though there are some relevant exceptional $A \wedge \neg B$-worlds; as long as such exceptions are not too widely spread, they will not have any semantic impact. In all of this, our chief interest will be in counterfactuals that are true or false in virtue of causal matters rather than, for example, by linguistic convention. Primarily it is such counterfactuals on which choosing our semantics over Lewis' will pay off, and it is such counterfactuals which come closest to hypotheses in empirical science.[6]

Summing up, we aim at

1. a scientifically plausible and simple probabilistic semantics for subjunctive conditionals,
2. which nevertheless turns out to stay quite close to the Lewis–Stalnaker semantics for conditionals,
3. such that the truth of $A \boxright B$ allows for exceptions, that is, where $\boxright$ comes with an implicit ceteris paribus clause.[7]

1, 2, and 3 are what we promise to have achieved by the end of part B of this paper, combined with the assurance that the semantics will build on, and in some ways extend, Adams' theory of conditionals, in a sense yet to be explained.

For now, assume the counterfactual

If the match were struck, it would light

to be true in spite of the minor possibility of exceptional circumstances. Where may these exceptions spring from? It might be that the world itself is chancy, so that there is a very small chance of the match not lighting if struck. It might also be that the antecedent is

---

[5] Although we will see in part B that this equivalence result will rest on a probabilistic version of the Limit Assumption that was criticized by Lewis, and although Lewis' standard of closeness does not necessarily coincide with the standard of closeness that is used in our theory.

[6] However, we will not have to say anything about causality itself in this paper, and in particular we do not want to analyze causality in terms of counterfactuals or vice versa; instead the semantics will remain purely probabilistic. For the same reason we will have to postpone a discussion of Judea Pearl's work on counterfactuals and causal reasoning (cf. Pearl, 1988, 1996) to a different occasion; but see Footnote 15 below. While our semantics is going to involve conditional chance, it will not involve anything like causal networks or at least it will not do so explicitly, unlike Pearl's theory.

[7] In other areas, tolerance with respect to exceptions has become a major topic quite some time ago: In Theoretical Computer Science, *default expressions* are investigated which allow for the existence of exceptions. In Philosophy of Science, *ceteris paribus* laws have been a long-standing subject of interest. In Metaethics, the study of prima facie obligations is prominent; and so forth. In particular, Schurz (2001, 2005) presents Adams' system of conditional logic as governing normic laws of the form 'if A, then normally B' in the life sciences; Smith (2007) treats ceteris paribus conditionals in a Lewis-style setting with normalcy components; and see van Benthem et al. (2009a) for a modal logic and semantics of ceteris paribus preferences.

imprecise so that it seriously underspecifies both the manner in which the match is struck and the relevant circumstances, which is why deviant but salient cases of striking without lighting cannot be excluded. Whether the source of these exceptions is, accordingly, ontic or semantic, we will propose that a probabilistic semantics for counterfactuals can—and should—take care of them. But our main emphasis will be on the chancy side.

Ideas similar to the ones that motivate the semantics in this article have been put forward by various authors: Skyrms (1980a, 1980b, 1984, 1994) and—building on Skyrms' work—Woodruff (1999) and Skyrms et al. (1999) tie the degrees of acceptability of counterfactuals to expected conditional chances. Adams (1975, 1976) and Edgington (1995, 2004, 2008) consider the idea that counterfactuals are epistemic past tense versions of indicative conditionals and hence may be treated in terms of conditional epistemic probability in ways similar to standard present tense indicatives. We will start with the work of Skyrms, Adams, and Edgington in Section §2 where we will explain how our new semantics will relate to such suppositional accounts of subjunctive conditionals. We are going to suggest that a counterfactual carries *both* a suppositional degree of acceptability *and* a truth-based degree of belief which may or may not coincide. So we will grant suppositionalists about counterfactuals their pragmatic notion of acceptance, however, we will nevertheless not side with them with regards to the thesis that counterfactuals are not truth-apt,[8] and we will not maintain that counterfactuals are anything like past tense indicative conditionals either. For instance, we take the match-striking conditional above to be formulated in the subjunctive mood, to be interpretable as describing an event which, if it were to happen at all, would happen now, and to express something about objective and graded possibilities, that is, ways the world might be like objectively. Apart from suppositional accounts of counterfactuals, Edgington (1995) cites (and criticizes) Blackburn (1986, pp. 213–255) and an unpublished manuscript by Michael Woods as presenting the truth condition for counterfactuals that we we are going to suggest in our own semantics. Bennett (2003, chap. 16) presents a truth condition similar to ours as being part of his "near miss" proposal which he regards as one option of saving Lewis' semantics for counterfactuals from yielding way too many unwanted falsities in the face of indeterminism. Kvart mentions the truth condition in which we are interested at various places (see Kvart, 1986, chap. 2; Kvart, 1992, footnote 2; Kvart, 2001, footnote 90; Kvart, unpublished), mostly in the context of discussions of probabilistic analyses of causality. Most recently, Loewer (2007) defends a truth condition such as the one of the semantics to be stated below when he relates the temporal asymmetry of counterfactuals to thermodynamical asymmetries that may be expressed in terms of statistical probabilities. Finally, Kaufmann (2005) studies so-called chance models of conditionals which derive from work done by Jeffrey and Stalnaker (cf. Stalnaker & Jeffrey, 1994); however, since according to these models conditionals do not always have classic truth-values at worlds, they are quite different from ours. We will return to various of these authors at various points of our article, and our theory should be seen as extending their work. But, to the best of our knowledge, our theory also goes beyond these references in various respects.

This is thus the plan of part A of the paper. In Section §2 we will reconsider Ramsey test accounts of conditionals in general, and of counterfactuals in particular. We will show how one may assign sensibly one semantic meaning and *two* pragmatic meanings to counterfactuals, and that these two permissible pragmatic understandings of counterfactuals

8 Woodruff (1999) is a special case here since he does consider a truth condition such as ours for counterfactuals, even though he aims to follow Skyrms' suppositionalist footsteps.

relate to two different kinds of degrees of acceptability or belief which are guaranteed to coincide if the world is deterministic, but which fall apart in the indeterministic case, as follows from a new triviality result on expected chance. However, we will also see that as far as the plain assertability of counterfactuals is concerned, the assertability conditions for counterfactuals that our theory predicts in terms of truth-based degrees of belief will normally be indistinguishable from the assertability conditions for counterfactuals that the suppositional theory predicts in terms of pragmatic degrees of acceptability. Hence, the pragmatics of counterfactuals that derives from our probabilistic semantics will turn out to be consistent with the suppositional theory of counterfactuals, as long as the latter is meant to govern only the acceptability of counterfactuals and not their believability-to-be-true. In Section §3 we will introduce the formal details of the new semantics which will be stated at first in terms of truth conditions for counterfactuals and then additionally in terms of their approximate truth conditions. Since the semantics is going to involve conditional probability measures, we will also enumerate the bits of probability theory that we will need in order to state the semantics in precise terms; moreover, we will present systems of conditional logic that are determined by both the truth part and the approximate truth part of the semantics. In Section §4 we will deal with the proper philosophical interpretation of our semantics: we will clarify its purpose, we will consider which type of probability it presupposes, we will deal with the time-relativity of the truth values of subjunctive conditionals, we will address some immediate worries that our semantics might cause, and we will discuss some of its advantages over Lewis' semantics. A short appendix with proofs concludes part A.

Part B of the paper will deal with the following topics: One of the differences between the new semantics and Lewis' will be that the so-called Centering Axioms[9] cease to count as logical truths. In section 1 of part B we will analyze why this is so, why it is not so bad, and what remnants of Centering the new semantics offers to the unconvinced. Along the way, we will also explain why and how our semantics coincides with Stalnaker's classic theory given the additional assumption of Counterfactual Determinism. In section 2 of part B we will suggest a second probabilistic semantics for counterfactuals. Although the second semantics will *look* quite different from the first one, it will follow from a new representation theorem for conditional probability measures that the two semantics are—in a sense explained by the theorem—*equivalent*. Since the metaphysical presuppositions of the second semantics will seem to be almost Lewisian, whereas the former semantics looks much more scientific, we will argue that in light of this representation result our probabilistic semantics may actually be understood as a naturalization of Lewis' semantics. Moreover, using the second semantics it will also be possible to explain why validity in the sense of our truth-conditional semantics coincides formally with validity according to Adams' non-truth-conditional theory on the restricted languages considered by Adams. Finally, section 3 in part B will consider an application of the new probabilistic semantics: As Hawthorne (2005) and Hájek (unpublished) have asked recently, are most ordinary counterfactuals false? We will see which answer is recommended by the new semantics.

§2. From the Ramsey test to the new semantics. Famously, in a footnote of his "General Propositions and Causality" (Ramsey, 1929), Frank Plumpton Ramsey proposed what is now called the Ramsey test for conditionals:

[9] That is, instances of: $(A \mathrel{\Box\!\!\rightarrow} B) \supset (A \supset B)$ and $(A \wedge B) \supset (A \mathrel{\Box\!\!\rightarrow} B)$.

> If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q... We can say that they are fixing their degrees of belief in q given p.

Stalnaker (1968, p. 102) generalized this to include cases in which the antecedent is believed to be false, as in many situations in which *counterfactuals* are assessed epistemically by an agent:

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

By itself, the Ramsey test is too weak and too imprecise to allow any particularly surprising conclusions to draw from it. This changes once it is positioned within a precise formal framework. There are, broadly speaking, two traditions of how do to do so, which differ with respect to the scale of measurement on which they analyze belief or acceptance: the qualitative theory of belief revision (cf. Levi, 1984, 1996; the AGM theory of Alchourrón et al., 1985; and Gärdenfors, 1988) which is interested in belief or acceptance *simpliciter*, and the quantitative theory of epistemic probability which measures *strengths* of belief or acceptance in terms of numerical degrees.[10] While the quote from Ramsey above falls squarely into the probabilistic tradition, it was taken up on the qualitative side, too, and the quote by Stalnaker above became extremely influential in both traditions from the start. We will only be concerned with the probabilistic approach to conditionals in this paper (parts A and B). But it should be noted that for most purposes it is in fact possible to translate theories and results within one tradition into corresponding theories and results in the other. In particular, there is a qualitative counterpart of Ernest Adams' probabilistic suppositional theory of conditionals, that is, Isaac Levi's theory of the suppositional acceptance of conditionals.

For *indicative* conditionals, such as the paradigmatic 'If Oswald didn't kill Kennedy, someone else did', Adams (1965, 1975) suggested to make the Ramsey test precise in probabilistic terms as follows:[11]

---

[10] There are also theories on scales of measurement that lie in between these two: see, for example, Spohn (1988).

[11] We leave all the formal details in this section, in which we will discuss suppositional accounts of conditionals and how they relate to our own semantic account, on a more or less intuitive level; for instance, we will not pause and treat cases with special care in which the relevant conditional probabilities would only be defined if primitive conditional chance functions were used; we will apply probability measures sometimes to sentences, sometimes to sets of possible worlds, and sometimes even to "conjunctions" of sentences and sets of worlds, where we will rather always apply them to propositions, that is, sets of worlds, later; and the like. Nor will we put forward in this section any background discussion on how terms such as 'conditional chance' are to be understood over and above their standard understanding in discussions on kinds of probability, we will leave the time relativity of chances implicit whenever we can, and so on. In contrast, when we turn to our own theory from the next section, we will make sure to be precise on all formal requirements—for example, we will explain the mathematics of primitive conditional probability measures at length—and we will deal with our intended interpretation of 'conditional chance' and the time-relativity of chances in detail.

- For every epistemic probability measure $\mathfrak{Cr}$, for all sentences A, B:

$$\text{The degree of acceptability } Acc_{\mathfrak{Cr}}(A \to B) \text{ (in } \mathfrak{Cr}\text{) equals } \mathfrak{Cr}_A(B),$$
$$\text{where: } \mathfrak{Cr}_A(B) = \mathfrak{Cr}(B|A).$$

So if an agent's epistemic state is given by a subjective—epistemic probability measure $\mathfrak{Cr}$, then the indicative conditional $A \to B$ is acceptable for the agent to a degree that coincides with the conditional probability of B given A as being determined by that measure. Here—as in the rest of this section—we denote by '$\mathfrak{Cr}(A)$' an agent's degree of belief that is assigned to the sentence A, however one should always think of this degree as actually being attached to the proposition that is expressed by A, that is, to the set of epistemically possible worlds in which A is true. On the other hand, if we write '$Acc_{\mathfrak{Cr}}(A \to B)$', we mean an agent's degree of acceptability of $A \to B$, where we will leave open at this point whether this degree actually derives from a degree that comes attached to a proposition. (We will use the same terminology in the subjunctive case further down below.) Finally, for the moment, think of A and B as not containing any nonmaterial conditional sign themselves, for it is mainly this case of simple conditionals $A \to B$ in which Adams was interested. Later in this section, when we compare suppositional and truth-based accounts of counterfactuals, we will have to say just a bit more about nested conditionals, and from Section §3 we will consider the full language of conditional logic, including arbitrary applications of propositional connectives to counterfactuals and arbitrary nestings of counterfactuals.

The idea behind Adams' formalization of the Ramsey test is of course that the antecedent of a conditional in the indicative mood is meant to apply to the actual world: *if A actually is the case*, then B holds. Accordingly, in the Ramsey test one should add the antecedent in the form of a *matter-of-fact* supposition (cf. Joyce, 1999); but a matter-of-fact supposition is very much like receiving evidence about the actual world for which conditionalization on the evidence is normally the appropriate and rational method of update, since what conditionalization does is just to cut down on the set of epistemically possible worlds by adding the evidence as an additional constraint. Moreover, in line with the Ramsey test, conditionalizing on A yields a new subjective probability measure $\mathfrak{Cr}_A$ which may then be used in order to evaluate the consequent. The only remaining problematic feature of Adams' formalization of the Ramsey test for indicative conditionals is that conditionalization does not yet capture the qualification 'hypothetically' in Stalnaker's quotation, because offline-supposition and online-learning are simply assimilated; fortunately, this does not matter much in typical cases in which both the antecedent and the consequent speak about physical states of affairs, rather than, for example, epistemic ones. We will be concerned only with such antecedents and consequents in this paper.[12]

Early on, Adams (1965, 1975) (and later Edgington; see, e.g., Edgington, 1995, pp. 238, 316f) suggested that *subjunctive* conditionals, such as the paradigmatic 'If Oswald hadn't killed Kennedy, someone else would have', might be treated as "epistemic past tense" (Adams, 1975, p. 103) indicative conditionals. If that were the right way to go, then one

[12] The difference may show up, however, whenever the antecedent and/or the consequent describe an agent's state of mind, as in the well-known Thomason conditionals such as 'If my wife is unfaithful, I don't know it'. In such cases the agent ought to assess a conditional by conditionalizing on the antecedent while additionally, and at the very same time, inhibiting any introspective access to her own state of mind.

could easily adapt the formalization of the Ramsey test above and turn it into an acceptability test for counterfactuals (ignoring all justified worries again about *hypothetically* adding the antecedent): while the agent's current credence function might be $\mathfrak{Cr}$, what the agent would have to do in order to determine her degree of acceptability for a counterfactual would be first to adopt a "prior" epistemic probability measure $\mathfrak{Cr}'$ that the agent would have held rationally, typically, shortly before the time of the event that is described the antecedent, and then to apply the probabilistic Ramsey test for indicative conditionals *relative to that measure* $\mathfrak{Cr}'$. The potential differences between $\mathfrak{Cr}$ and $\mathfrak{Cr}'$ might thus explain the diverging evaluations of indicative and subjunctive conditionals as exemplified by the two famous Oswald–Kennedy conditionals: go back in time just before the shooting; back then you would not have accepted the indicative 'If Oswald does not kill Kennedy, someone else will', which is why *now* you do not accept the subjunctive 'If Oswald had not killed Kennedy, someone else would have'; at the same time, you do *now* accept the indicative 'If Oswald did not kill Kennedy, someone else did'.

Unfortunately, as Adams noticed himself, this strategy will not do as a *general* strategy (this issue is also taken up by Edgington, 1995, p. 318f). He discusses the following counterexample in Adams (1975, p. 129):[13]

> Imagine the following situation. We have just entered a room and are standing in front of a metal box with two buttons marked '*a*' and '*b*' and a light, which is off at the moment, on its front panel. Concerning the light we know the following. It may go on in one minute, and whether it does or not depends on what combinations of buttons *a* and *b*, if either, have been pushed a short while before, prior to our entering the room. If exactly one of the two buttons has been pushed then the light will go on, but if either both buttons or neither button has been pushed then it will stay off. We think it highly unlikely that either button has been pushed, but if either or both were pushed then they were pushed independently, the chances of *a*'s having been pushed being 1 in a thousand, while the chances of *b*'s having been pushed is a very remote 1 in a million. In the circumstances we think that there is only a very small chance of 1,000,999 in one billion (about 1 in a thousand) that the light will go on, but a high probability of 999 in a thousand that *if b was pushed, the light will go on*.
>
> Now suppose that to our surprise the light does go on, and consider what we would infer in consequence. Leaving out numerical probabilities for the moment, we would no doubt conclude that the light probably lit because *a* was pushed and *b* wasn't, and not because *b* was pushed and *a* wasn't. Therefore, since *a* was probably the button pushed, *if b had been pushed the light wouldn't have gone on,* for then both buttons would have been pushed. The point here is that the counterfactual would be affirmed *a posteriori* in spite of the fact that the corresponding indicative was very improbable *a priori*, because its contrary "if *b* was pushed then the light will go on" had a probability of 0.999 *a priori*.

[13] We replace Adams' original symbols '*A*' and '*B*' by '*a*' and '*b*', respectively, in this quote, in order to avoid a clash with our standard sentential metavariables.

What is going wrong in the past indicatives account is that when one evaluates a counterfactual such as in the story above, one is not so much drawn to some *temporally prior* epistemic probabilities but rather to some different *kind* of probabilities. In Adams' (1976, p. 17) words, when commenting on a similar urn-drawing example: "the identification of the 'prior' conditional probability, which is a counterfactual's probability, with the proportion of black balls in Urn A (whether or not this proportion is known before a ball is actually drawn) suggests that it might be somewhat less misleading to employ the occasionally used term 'propensity probabilities' in application to them." Accordingly, in the example from before, the agent should not switch to an epistemic probability measure *prior* to the pushing of the button in order to evaluate the counterfactual, but she should estimate the *conditional chance* of its consequent given its antecedent from her current point of view, that is, at the time when she is evaluating the counterfactual and when she had seen that the light had gone on. Adams (1975, p. 130) adds that it is "dubious that this counterfactuals's probability can be assumed to satisfy the usual laws of conditional probability." Instead, on pp. 131f he suggests what he calls an "*ad hoc*" model in which he calculates a counterfactual's probability in terms of a weighted sum of probabilities of causal or nomological connections relating the antecedent of the counterfactual together with a physical state to the consequent of the counterfactual, where the weights are the "posterior" epistemic probabilities of the states. This is exactly what was proposed and worked out in detail not much later by Skyrms (1980a, 1980b, 1984, 1994) who gives the following systematic analysis of the acceptability of counterfactuals in terms of expected conditional chance or propensity. If presented in the form of a probabilistic Ramsey test for counterfactuals (and adapting Skyrms' original terminology just a bit), it goes like this:

- For every epistemic probability measure $\mathfrak{Cr}$, for all chance measures $\mathfrak{P}_w$ in worlds w, for all sentences A, B:

$$\text{The degree of acceptability } Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B) \text{ (in } \mathfrak{Cr}\text{) equals } \mathfrak{Cr}^*_A(B),$$

$$\text{where: } \mathfrak{Cr}^*_A(B) = \sum_w \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A).$$

Hence, if an agent's epistemic state is given by an epistemic probability measure $\mathfrak{Cr}$ again, then the subjunctive conditional $A \mathrel{\Box\!\!\rightarrow} B$ is acceptable for the agent to a degree that coincides with the expected conditional chance of B given A as being determined by that measure. Here Skyrms assumes that for each epistemically possible world w there is a uniquely determined and well-defined single case conditional chance $\mathfrak{P}_w(B|A)$ of B given A relative to that world, and counterfactuals are accepted in terms of "averaging" over all of them according to the epistemic probabilities that one assigns to these worlds. He also presupposes that the epistemic probability measure is defined on an atomic algebra which includes every singleton set of the form $\{w\}$—or a state description sentence for every such singleton—where w is any world in the given set W of epistemically possible worlds. Additionally, we assume that $\mathfrak{Cr}$ is a standard absolute probability measure such that for every sentence C, $\mathfrak{Cr}(C)$ equals the sum of all $\mathfrak{Cr}(\{w\})$ over those worlds w which satisfy C, whether there are infinitely many such worlds w or not. We are going to abide by these simplifying assumptions throughout this section, but we will not do so beyond, since they are ultimately much too restrictive. In the current context, this won't matter really: in particular, all of these assumptions are satisfied, of course, when W is finite, which for most practical purposes is plausible to assume for *epistemic* probability measures over *epistemically* possible worlds anyway. In contrast, from the next section we will deal with the conditions under which counterfactuals are made true by the conditional *chance* measures of *physically* or perhaps *metaphysically* possible worlds, which at least *prima facie* does not involve any reference to epistemic probability measures at all, and for which, for example, the assumption of just a finite set W of physically or metaphysically set would be utterly implausible.[14]

The guiding thought behind Skyrms' formalization of the Ramsey test for counterfactuals may be reconstructed as follows: The antecedent of a conditional in the subjunctive mood is not meant to apply to the actual world, but rather if the actual world *were changed in the way such that A would be the case*, then B would hold. Accordingly, in the Ramsey test one should add the antecedent to one's beliefs in the form of a *contrary-to-the-facts* supposition (cf. Joyce, 1999) or maybe by means of an *open* or *not-necessarily-matter-of-fact* supposition; the supposition is not intended at all to change the agent's view of what the actual world is like, instead the agent "zooms into" her more or less epistemically likely candidates for the actual world, determines within each of them their worldly chances of B given A, and sums up the results as being weighted by their epistemic probabilities. Although Skyrms himself does not make this point, one may present his account—as we did above—in terms of an updated credence function Cr_A^* which then gets applied to the consequent B in question: the subjunctive supposition of the antecedent thus yields a new epistemic probability measure on the basis of which one may argue about the consequent, in line with the Ramsey test format.[15] For instance, in Adams' example above, the epistemically possible world(s) in which a and only a got pushed dominate subjectively at the time of the appraisal of the counterfactual, since at the time the agent already knows that the light went on (and a being pushed is by far the most likely explanation); in those worlds the chance of the light going on given that b had been pushed back then is zero, which is why overall 'if b had been pushed then the light would have gone on' is rejected by the agent. Note that $\mathfrak{Cr}_A^*$ may differ significantly from the conditionalized measure $\mathfrak{Cr}_A$ from before; accordingly, 'if b has been pushed then the light has gone on' is of course assigned a degree of acceptability very close to 1 at the time when the agent knows that the light has in fact

---

14 Additionally, one would actually need to specify *relative to what temporal reference point* the conditional chance $\mathfrak{P}_w(B|A)$ at w ought to be taken in Skyrms' formula. Skyrms' terminology of "prior" conditional propensities which he uses at some places is modelled after Adams' "prior" conditional epistemic probabilities, but it is also a bit misleading: *any* temporal reference point could be chosen in principle and sometimes more than one might be pragmatically salient; Skyrms (1980, p. 261) shows that Skyrms is well aware of that. We will postpone a general discussion of the time-relativity of chances to Section 4.3. where we will deal with this in detail. However, the short answer in the current context is: Normally, the degree of acceptability $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B)$ at time t_r is determined by the agent's credence function $\mathfrak{Cr}$ at t_r *and by the conditional chances* $\mathfrak{P}_w(B|A)$ *at the same time* t_r, independent of whether A or B speak about events at times other than t_r. Furthermore, when a counterfactual is asserted, it is normally the time of assertion which functions at the temporal reference point so that it is the conditional chances *at that time* which are meant to be taken. See the third example of Section 4.7 and in particular Footnote 72 in that section for an illustration.

15 Skyrms does not do so himself, but one can prove easily that Cr_A^* is in fact an absolute probability measure given that Cr is an absolute probability measure and each $\mathfrak{P}_w$ is a Popper function for which the proposition expressed by A is not abnormal, in the sense of Section §3. An update operation such as Cr_A^* also figures prominently in Judea Pearl's action-based transformation of states based on transition probabilities, with Pearl's transition probabilities being Skyrms' conditional chances; see, for example, Pearl (2000, sec. 2.5). Pearl (1994, p. 71) refers to Skyrms' work on this matter.

gone on. On the other hand, generally speaking, whenever an agent knows the conditional chance of B given A (and hence it is the same in all epistemically possible worlds), and the agent does not have any inadmissible knowledge (cf. Lewis, 1980), then $\mathfrak{Cr}^*_A(B)$ will be precisely that conditional chance which in turn may be expected to coincide with $\mathfrak{Cr}_A(B)$, by a conditional variant of Lewis' Principal Principle (see section 1.3 in part B; a similar point was observed already by Skyrms). Furthermore, in many cases in which an agent does not have any relevant evidence concerning the truth value of either the antecedent or the consequent of a subjunctive conditional at all, and when the agent does not know anything about the causal structure in question, $\mathfrak{Cr}^*_A(B)$ may be expected to be close to or even identical to $\mathfrak{Cr}_A(B)$, by some reflection principle again. In particular, this should be the case often when A describes a future event from the viewpoint of the evaluating agent: No wonder we generally assign the same degrees of acceptability to future indicatives and subjunctives! Time and tense only play a role here insofar as conditionals are of course always evaluated at a time, and also conditional chances are always relative to points of time (which we will discuss in detail in Section 4.3 when we consider the role of chance in our own semantics). Otherwise there is nothing at all about subjunctive conditionals which would tie them to any particular tense—such as past tense, as the past tense indicative view of counterfactuals had it—and indeed subjunctive conditionals may speak about present or future events just as much as indicative conditionals may do.[16]

We submit that the suppositional theory of conditionals—including counterfactuals, on which we will now focus, in particular—yields an excellent pragmatic theory of conditionals. This said, we also maintain that as a theory of counterfactuals, indeed even as a theory of the *pragmatics* of counterfactuals, it is incomplete. In order to see why, let us reconsider the subjunctive conditional

(1) If the match were struck, it would light

from the last section. As step one, it should be unproblematic to qualify the consequent of (1) in either of the following ways, while leaving the meaning of the conditional—completely or at least almost completely—unaffected:

(2) If the match were struck, it would necessarily [definitely] light.

(3) If the match were struck, it would be very likely to light.

Without any bias from some philosophical theory, (2) and (3) simply seem to say pretty much the same as (1). Either we regard (1) as taking the lighting of the match to follow its being struck invariably, or one is more tolerant about the potential existence of exceptional circumstances: in the former case, (2) ought to be synonymous to (1), while in the latter

16 Edgington (1995, p. 319) does not regard Adams' example as a proper counterexample to the past tense indicatives view of counterfactuals: she suggests that evaluating the counterfactual in question would be like evaluating the indicative conditional antecedently given one would have known also that b had been pushed. Instead, the example "brings out some of the subtleties involved in trying to assess objective chances," subtleties which, as she does cite, Skyrms addresses in his theory. But that is certainly an understatement: in order for the past tense indicatives theory to go through, one would need a systematic account of when to include some information on the antecedent or the consequent within the "prior" epistemic probability measure that is for the agent to be used to assess a counterfactual (as would be necessary for Adams' button-pushing example) and when such information should be disregarded (as in Adams' Oswald–Kennedy example). Skyrms' theory offers a systematic treatment of both cases—and many more, see Skyrms (1984, pp. 261–263)—but crucially his is *not* a past tense indicatives account.

case, (3) would be synonymous to (1). But even in the former case, (3) would be "almost" synonymous with (1), and for many purposes the difference between (2) and (3) will not play any major role anyway. For that reason, but of course also because conditionals such as (3) will be the main topic of the rest of this paper, and because the probabilistic rendering of (1) makes it easier to relate the current discussion to the suppositional views from above which have been probabilistic from the start, we will focus just on (3) as being synonymous or nearly synonymous to (1).[17]

Next, we intend to analyze (3) with the aim of reformulating it somewhat more transparently. A sentence such as (3) invites two obvious questions: Which kind of probability is expressed by the qualification 'very likely'? And what is the logical form of the whole conditional statement? Without running through all the options that seem possible at first glance, the following answers seem to be the most salient ones by far: The probability in question is an ontic one—single-case chance again—and the logical form of (3) is the one of:

(4) The conditional chance of the match lighting given that it is struck is very high.

After all, (3) seems to be true or false depending only on whether the match is wet or not, whether there is enough oxygen, and so on, in short: depending only on what our physical world is like. In a different possible world, circumstances might be different, the chance of the very same match lighting if struck might not be high anymore, and the truth value of (3) might thus differ accordingly.[18] (3) does not seem to say anything epistemic, at least not in any direct or obvious manner: it is not some agent's state of mind which is making (3) true.[19] Indeed, (3) ascribes to the particular match in question a particular physical property, that is, the graded disposition or tendency of flammability.[20] Quantum theory or maybe even statistical mechanics (cf. Section 4.4) might in principle give us a way of quantifying this disposition or tendency in terms of conditional chance, but sometimes we can estimate such numbers quite reliably just by experience and symmetry considerations. In any case, the probability in question should be an objective nonepistemic one. And as far as the logical form of (3) is concerned, it is difficult to see that any logical analysis other than in terms of a conditional probability ascription—whether one would try a combination of some 'if-then' connective with an absolute probability operator in the consequent, or the application of an absolute probability function sign to a proposition that gets expressed by

---

[17] Of course, we do not claim that 'necessarily' and 'very likely' are strictly synonymous. Even 'maximally likely' (i.e., having probability 1) and 'necessarily' differ in meaning.

[18] We presuppose that one may identify physical objects across worlds, that one can refer to such objects by means of rigid designators, and that 'the match' picks out an object and refers to that object just as a rigid designator does—but similar assumptions are made by most modern modal semantics anyway.

[19] The terms 'direct' and 'obvious' are meant to pacify those who hold a theory of objective chance according to which chances are ultimately nothing but projections of properties of an ideal observer's mind onto the world. Even proponents of such a view will normally grant others their using the term 'objective chance', however, they will tacitly substitute their own favorite subjectivist or empiricist rendering of that term whenever necessary.

[20] We do not claim that dispositions such as flammability can be *defined* in terms of subjunctive conditionals, as the traditional conditional analysis of dispositions had it which in recent years has come under serious attack. We only claim that in the case of a normal match in normal circumstances, (3) above is made either true or false by the presence or the absence of a dispositional property of the match, which seems hard to deny.

some sort of conditional, or the like— could be equally faithful to what we normally intend to say with (3).

About the question of logical form: Barker (1999) considers all typical attempts of how to analyze (3) in terms of a Lewisian $\Box\!\!\rightarrow$ connective and/or a Lewisian notion of closeness, joined in some way with a probabilistic operator, and he presents arguments against all of them; this eventually leads him to introduce his own causal algorithmic theory of probabilistic counterfactuals. We will not recapitulate his arguments here, but, for example, he rules out successfully the understanding of (3) in terms of a Lewisian counterfactual with a probabilistic operator in the consequent but none in the antecedent.[21] The account that comes closest to (4) above among those which he discusses is in terms of a proportion statement of the form 'ω of the closest P-worlds are Q-worlds' (using his symbols and terminology, see Barker, 1999, p. 433). His criticisms of it are based on the vagueness of Lewisian similarity, the problem of defining proportions over infinite sets, and some features of Lewis' semantics for nonprobabilified counterfactuals, none of which apply in our own context. In section 2.5 of part B we will show how one can in fact represent statements such as (4) equivalently in terms of a Lewisian $\Box\!\!\rightarrow$ and probabilistic operators, but we will also see that this representation is only possible subject to certain assumptions, that it needs an absolute probability ascription in the antecedent *and a conditional probability ascription in the consequent*, that the whole representation is much less perspicuous than (4) above, and that it is none that Barker considers. In short, (4) stands its ground.[22]

So it seems we have hit upon a very reasonable (almost) reconstruction of

(1) If the match were struck, it would light

by means of

(4) The conditional chance of the match lighting given that it is struck is very high.

But this should leave us puzzled as how to reconcile this fact with the plausibility of the suppositional story as being told before: for by Lewis' celebrated triviality or impossibility theorems (which we will discuss in detail in Section 4.5), conditional probabilities cannot generally and in any straightforward manner (short of abandoning classical logic or introducing hidden epistemic contextual parameters or the like) be determined *qua* absolute probabilities of conditional propositions. Thus, the suppositional degree of acceptability that is assigned to an indicative conditional cannot be equated with the absolute epistemic

[21] Here is an argument different from Barker's against an analysis of (3) in terms of a counterfactual with just an absolute probability operator in the consequent: if this were a faithful analysis, then accordingly it should be possible to represent the true 'if the (fair) die rolled a 1 or a 2 or a 3 or a 4, the probability for it to roll a 1 would be $\frac{1}{4}$' by a statement of the form '$(1 \vee 2 \vee 3 \vee 4) \Box\!\!\rightarrow P(1) = \frac{1}{4}$'; and similarly the true 'if the (fair) die rolled a 1 or a 4 or a 5 or a 6, the probability for it to roll a 1 would be $\frac{1}{4}$' should be representable by a statement of the form '$(1 \vee 4 \vee 5 \vee 6) \Box\!\!\rightarrow P(1) = \frac{1}{4}$'. But by the rule of Disjunction, which is valid according to all standard logics for conditionals, these two representations would then logically entail '$(1 \vee 2 \vee 3 \vee 4 \vee 5 \vee 6) \Box\!\!\rightarrow P(1) = \frac{1}{4}$' which would represent the false 'if the (fair) die rolled a 1 or a 2 or a 3 or a 4 or a 5 or a 6, the probability for it to roll a 1 would be $\frac{1}{4}$'.

[22] Barker (1999, p. 435f) also argues briefly against both Edgington's and Skyrms' suppositional treatments of the acceptability of counterfactuals, based on examples of the same kind as the plane crash example that we are going to discuss in detail in Section 4.7 in the context of our own semantics. What we will say there can be translated directly into a defense of Skyrms' suppositional theory against Barker's criticism; see in particular Footnote 72 in that section.

probability of a proposition, that is, one cannot have for all epistemic probability measures (whether "prior" or determined from a "prior" one by conditionalization) on a language that includes $A \to B$ and which is closed under the standard operators of propositional logic that

$$Acc_{\mathfrak{Cr}}(A \to B) = \mathfrak{Cr}(A \to B).$$

But the same must be true of the suppositional degree of acceptability that is assigned to a counterfactual conditional, for otherwise expected conditional chances would have to be equal to absolute epistemic probabilities also in those quite common cases in which expected conditional chances actually coincide with the corresponding epistemic conditional probabilities—but that has been ruled out before by Lewis' theorems. So the suppositional degrees of acceptability for counterfactuals could not in any straightforward way be degrees of belief of propositions being true, which they should be—or so it seems—if (4) is a good reconstruction of (1). After all, (4) does in fact expresses a proposition that is true or false at worlds in the standard sense and hence believing it partially should amount to assigning it an absolute epistemic probability in the standard sense, whereas suppositionalists usually take the considerations from Lewis' results to show that conditionals do not express propositions and thus not to have classical truth values at all, which is also why they cannot be assigned an absolute epistemic probability in the usual sense. Hence the puzzlement.

Here is the precise argument why degrees of acceptability in the sense of Skyrms cannot generally be degrees of belief for propositions in the standard sense: For assume that counterfactuals express propositions and are therefore true or false at worlds, and assume also that all credence functions to be considered assign to each counterfactual a degree of belief that derives from the proposition that is expressed by the counterfactual and which coincides with the corresponding degree of acceptability as computed by the corresponding expected conditional chance. Now consider a counterfactual $A \mathrel{\Box\!\!\rightarrow} B$ and an epistemically possible world w' with an epistemic probability $\mathfrak{Cr}(\{w'\}) > 0$. We write '$(A \mathrel{\Box\!\!\rightarrow} B) \wedge \{w'\}$' for the conjunctive proposition that is formed by the intersection of the proposition expressed by $A \mathrel{\Box\!\!\rightarrow} B$ and the singleton proposition $\{w'\}$. It follows:

$$\begin{aligned}
&\mathfrak{Cr}((A \mathrel{\Box\!\!\rightarrow} B) \wedge \{w'\}) = && \\
&= \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B \,|\, \{w'\}) && \text{(Ratio formula, conditional prob.)} \\
&= \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{Cr}_{\{w'\}}(A \mathrel{\Box\!\!\rightarrow} B) && \text{(Def. of } \mathfrak{Cr}_A\text{, p. 32)} \\
&= \mathfrak{Cr}(\{w'\}) \cdot Acc_{\mathfrak{Cr}_{\{w'\}}}(A \mathrel{\Box\!\!\rightarrow} B) && \text{(Degree of Acc. = Degree of Belief)} \\
&= \mathfrak{Cr}(\{w'\}) \cdot [\mathfrak{Cr}_{\{w'\}}]^*_A(B) && \text{(Degree of Acc., p. 34, applied to } \mathfrak{Cr}_{\{w'\}}\text{)} \\
&= \mathfrak{Cr}(\{w'\}) \cdot \textstyle\sum_w [\mathfrak{Cr}_{\{w'\}}](\{w\}) \cdot \mathfrak{P}_w(B|A) && \text{(Def. of } \mathfrak{Cr}^*_A\text{, p. 34)} \\
&= \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{P}_{w'}(B|A). && \text{(Calculate)}
\end{aligned}$$

Here $\mathfrak{Cr}$ gets conditionalized on $\{w'\}$, and the resulting epistemic probability measure is then plugged into Skyrms' definition of the degree of acceptability of $A \mathrel{\Box\!\!\rightarrow} B$.

Now there are just two possible cases:

- $A \mathrel{\Box\!\!\rightarrow} B$ is true in w': In this case we may conclude from

$$\mathfrak{Cr}(\{w'\}) = \mathfrak{Cr}((A \mathrel{\Box\!\!\rightarrow} B) \wedge \{w'\}) = \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{P}_{w'}(B|A)$$

and $\mathfrak{Cr}(\{w'\}) > 0$ that

$$\mathfrak{P}_{w'}(B|A) = 1.$$

- $A \,\Box\!\!\rightarrow B$ is false in w': In this case we have from

$$0 = \mathfrak{Cr}(\varnothing) = \mathfrak{Cr}((A \,\Box\!\!\rightarrow B) \wedge \{w'\}) = \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{P}_{w'}(B|A)$$

and $\mathfrak{Cr}(\{w'\}) > 0$ that

$$\mathfrak{P}_{w'}(B|A) = 0.$$

This shows: Assume a counterfactual $A \,\Box\!\!\rightarrow B$ expresses a proposition and thus has a classical truth value at w'. If the Skyrmsian degree of acceptability of the counterfactual as being determined by a credence function is taken to be equal to the degree of belief that is assigned to that proposition under that credence function, that is,

$$Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) = \mathfrak{Cr}(A \,\Box\!\!\rightarrow B)$$

for all epistemic probability measures $\mathfrak{Cr}$, then given the assumption $\mathfrak{Cr}(\{w'\}) > 0$ for any particular credence function $\mathfrak{Cr}$, it follows that the conditional chance of B given A in w' must be either 1 or 0, and indeed equal to 1 if and only if $A \,\Box\!\!\rightarrow B$ is true in w'. In a nutshell: Apart from worlds which have been ruled out epistemically, the suppositional degree of acceptability for $A \,\Box\!\!\rightarrow B$ is equal to the degree of belief of $A \,\Box\!\!\rightarrow B$ being true *only if conditional chances are deterministically crisp and the truth of a counterfactual consists in its corresponding conditional chance being 1 ("maximally high")*. This is the triviality result on conditional chance and counterfactuals which is the counterpart of Lewis' for conditional epistemic probability and indicative conditionals; conditional chances get trivialized in the given circumstances in the sense that they are bound to be 1 or 0.[23]

Furthermore, and in turn, *if conditional chances are deterministically crisp in every world, and if a counterfactual $A \,\Box\!\!\rightarrow B$ is true at a world if and only if the conditional chance of B given A at that world is 1 ("maximally high")*, then the suppositional degree of acceptability for $A \,\Box\!\!\rightarrow B$ is equal to the degree of belief in the truth of $A \,\Box\!\!\rightarrow B$ relative to any epistemic probability measures $\mathfrak{Cr}$, since then

$$Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) = \mathfrak{Cr}^*_A(B) = \sum_{w} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) =$$

$$\sum_{w \models A \Box\!\rightarrow B} \mathfrak{Cr}(\{w\}) \cdot 1 + \sum_{w \not\models A \Box\!\rightarrow B} \mathfrak{Cr}(\{w\}) \cdot 0 \;=\; \mathfrak{Cr}(A \,\Box\!\!\rightarrow B).$$

We may draw two conclusions from this: First of all, in the deterministic case, our reconstruction of (1) in terms of (4) from above does not contradict the suppositional theory of acceptance of counterfactuals; in fact, it is the only plausible propositional construal of counterfactuals that leads to the same pragmatic predictions as the suppositional theory. For one needs truth conditions for $A \,\Box\!\!\rightarrow B$ then which are such that it follows that $A \,\Box\!\!\rightarrow B$ is true in w if and only if the conditional chance of B given A is 1 at w. In order to guarantee this, one needs to define truth in terms of conditional chance. One way to do so is to define truth for $A \,\Box\!\!\rightarrow B$ at w in terms of '$\mathfrak{P}_w(B|A) = 1$', that is, $\mathfrak{P}_w(B|A)$ being "maximally high." Another possibility is to define truth by means of

[23] We use the predicate 'deterministic' here simply as a shorthand label for: all world-relative conditional chance values are either 1 and 0. Of course, this condition actually allows for various interpretations. See Section 4.4 for a discussion on this.

'$\mathfrak{P}_w(B|A) \geq 1 - \alpha$' for some $\frac{1}{2} < \alpha \leq 1$, that is when $\mathfrak{P}_w(B|A)$ is just "high" or "very high"; this won't make a difference, since by determinism all conditional chances at worlds are either equal to 1 or 0 anyway. Indeed, it would even work to define truth like that for any $0 < \alpha < \frac{1}{2}$, or even straightforwardly by means of '$\mathfrak{P}_w(B|A) > 0$', however neither of these two options would lead to anything that we would still recognize as being sufficiently synonymous to the truth conditions of subjunctive conditionals as we understand them ordinarily. So the only plausible definition of truth for counterfactuals that is consistent with the suppositional theory of counterfactuals, if given determinism, is in terms of "maximally high" or just "high" conditional chance. In Section §3 we are going to introduce the corresponding notions of truth for counterfactuals which we will then call 'truth (*simpliciter*)' and 'approximate truth (to degree $1 - \alpha$)'.[24]

Secondly, in the indeterministic case, if there are worlds with positive epistemic probability which allow for intermediate conditional chances that lie strictly between 0 and 1 as measured at those worlds, our reconstruction of (1) in terms of (4) *does* seem to contradict the suppositional theory of the acceptance of counterfactuals. So the intuitive puzzlement from before is confirmed on theoretical grounds, if "only" in an indeterministic world (such as, presumably, our own one).[25]

One possible reaction to this predicament is for the suppositionalist to bite the bullet and to claim the considerations from before, concerning the alleged (almost) synonymy of (1) and (4), to be misguided in some way: maybe by some pretheoretic and tacit adherence to the unproblematic deterministic case, maybe by falsely believing that (4) is made true or false by some dispositional property in the world, or by some other flaw. Indeed, Adams himself (cf. Adams, 1976, footnote 15) seems to toy with the idea of going antirealist, when he says:[26]

> That counterfactual's probabilities might be identified in certain cases with the probabilities of dispositional propositions calls into question our working hypothesis that truth values won't apply to counterfactuals. At any rate, we have no hesitation in analyzing dispositionals as though they were truth-conditional in investigating their deductive interrelations

[24] Additionally, in Section §3 we also suggest a third interpretation of 'very high conditional chance' as a vague term. This interpretation will be defended in section 2.3 of part B on the basis of work done by Adams. However, we will not deal with this interpretation at all in the current Section 2.

[25] Two remarks: Firstly, we learned in personal communication that also Robbie Williams and Rachael Briggs are working on triviality results for counterfactuals. Secondly, in section 1 and 2 of part B we will find that if all conditional chance functions in worlds are assumed to be counterfactually deterministic in the sense that they only take crisp values in the set (rather than the interval) $\{0, 1\}$, independent of the choice of antecedent or consequent, then the resulting logic of counterfactuals will be Stalnaker's, and our world-relative conditional chance functions will then correspond to Stalnakerian world-relative selection functions. In that case, the epistemic operation $\mathfrak{Cr}^*_A$ is nothing but Lewis' well-known imagining operation, which he suggested in Lewis (1976). In this sense, Skyrms' update operation is a generalization of imaging to the indeterministic case.

[26] Levi (1977) does the same in his qualitative suppositional account of counterfactuals and disposition terms in terms of contractions and expansions of a corpus of knowledge; as he says on p. 435, "Either we should reject efforts to give truth conditions for disposition statements in modal terms or we should embrace a more realistic construal of modal statements and move down the primrose path of modal semantics realistically interpreted. Needless to say, I favor the former alternative."

> and we also attach non-conditional probabilities to them. The theoretical issues involved here have some connections with the so called problem of *realism* of dispositional statements...

But wouldn't it be quite surprising if the puzzlement from above—which resulted from a simple semantic and pragmatic analysis of counterfactuals—could be resolved only by taking sides in an important and representative portion of the age-old debate of realism versus antirealism?

And even if that were so, other serious issues would still remain. We will deal with two of them. First of all, consider a *negated* counterfactual such as the negation of (1),

(5) It is not the case that if the match were struck, it would light

which seems to be straightforwardly understandable. Indeed, if (1) may be understood in terms of (4), then (5) has an obvious reading:

(6) It is not the case that the conditional chance of the match lighting given that it is struck is very high.

Accordingly, one ought to believe (5) by the same degree of belief that one assigns to the truth of (6), which should be 1 minus the degree of belief for (4). So far, so good.

Now, how is the suppositional theory to make sense of (5), and which degree of acceptability would it have to assign to it? One option would be to claim that (5) denies that the expected conditional chance of the match lighting given struck is very high: but if so then (5) denies a proposition about an epistemic state of affairs—epistemic because 'expected' really means: 'expected-relative-to-$\mathfrak{Cr}$'—even though (1) was not supposed to express a proposition at all, and in spite of (5) not seeming to be speak about anyone's state of mind at all. Semantically (or pragmatically?), 'it is not the case that' would thus not so much take a proposition as its argument in order to map that proposition to another one, but it would have to turn something else—a mental state? an aggregate of a syntactic item and a degree of acceptability?—into a proposition; that proposition could then be assigned a degree of belief in the standard way. These consequences could be avoided if already (1) had been taken to express a proposition about someone's epistemic state, that is, to express the proposition that someone's expected conditional chance of the consequent of (1) given the antecedent of (1) would be very high. But for the same reason as in the indicative case, suppositionalists generally reject this strategy: for them, conditionals are meant to express high epistemic conditional probabilities or high expected conditional chances but not *that* these numbers were high—otherwise the subject matter of (1) would be someone's mind rather than the world, in contrast with how we normally understand (1), and the suppositionalist would not be in a better position to deal with the triviality results than the nonsuppositionalists are (see Bennett, 2003, sec. 42 for details on this "suppositional expressivism").

The previous line on negated counterfactuals was problematic for the suppositional theory in view of the fact that while there are negations of propositions about some probability being high, it is hard to make sense of negated high conditional probabilities. However, maybe one can in fact make sense of the latter at least numerically in terms of 1 minus the respective expected conditional chance. Accordingly, the other take on (5) that a defender of the suppositional theory might put forward would be to supply (5) with the degree of acceptability that is given by $Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} \neg B)$, where A and B are the antecedent and the consequent of (1), respectively—call this the suppositional option 2 concerning negated counterfactuals. Presumably, this could only be justified on grounds of interpreting (5) as

(7) If the match were struck, it would not light.

For the sake of the argument, let us suppose that this is acceptable. If it is, then it should be possible for the suppositionalist to apply the same strategy to

(8) It is not the case that if the match were struck, it would be very likely to light.

We are not assuming anymore that (1) and (3) are (almost) synonymous, as this is under dispute, instead (3) is taken to be a probabilistic counterfactual that is of interest in itself, and (8) is simply its negation. Ignoring any moves which would themselves be problematic again as discussed before—such as taking (3) to be a standard counterfactual with a probabilistic operator in the consequent, or thinking of (3) as expressing a proposition about some person's epistemic state—there is in fact a truly suppositional treatment of (3) according to which the degree of acceptability for it should be $Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$ again or maybe *some number slightly higher than* the original $Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$ (since, if anything, it should be slightly easier to accept (3) than (1), independent of the question of whether (3) is almost synonymous to (1)). This is vague, but let us put that aside for the moment. Indeed, a suppositional account of (3) should be able to simply take 'if. . . , then very likely . . . ' as another if-then connective to which some version of the Ramsey test applies, and that version should leave (3) with a degree of acceptability very much like that of (1). But if so, the alleged analysis of (5) in terms of (7) should then lend support to the corresponding analysis of (8) in terms of

(9) If the match were struck, it would be very likely not to light.

But it seems the proper understanding of (8) should really be

(10) If the match were struck, it would not be very likely to light,

since in the case of a match that someone might be about to strike and which we know has a chance of 0.5 of lighting, we would happily claim (8) and (10) with equal confidence while at the same time rejecting (9) quite confidently, too. But since that has gone wrong, the suppositional option 2 of how to determine a degree of acceptability for (5) by effectively understanding (5) in terms of (7) should be questionable as well. And the only way of avoiding this, it seems, would be to return to analyzing (3) in terms of a probabilistic operator in the consequent, such that the negation sign would end up being outside of the probabilistic context, as in (10), rather than inside, as in (9). But analyzing (3) in that way is not very promising, as we pointed out before. Hence, neither of the two suppositional options concerning negated counterfactuals really looks promising either.

The situation even gets worse if counterfactuals with *counterfactual antecedents* are considered (this is our second problem case for suppositionalists after the one for negated counterfactuals):

(11) If it were the case that if the match were struck, it would light, then I could change that easily by dipping the match into water

is perfectly understandable again, and it does not cause any problems whatsoever for the propositional reading of counterfactuals that was proposed above either: (11) is understood as being (nearly) synonymous with

(12) If it were the case that the conditional chance of the match lighting given that it is struck is very high, then I could change that easily by dipping the match into water

which in turn may be taken to say (more or less)

(13) Given it is the case that the conditional chance of the match lighting given that it is struck is very high: then the conditional chance of me changing that easily by dipping the match into water is very high on that condition.

To be sure, this involves a nesting of propositions about conditional chance, which might not be to everybody's taste, but it is also not particularly worrisome once chances have been relativized to worlds (as Skyrms himself has it). The degree of belief for (13) is then nothing but the degree of belief assigned to the proposition that is expressed by (13), as always in a propositionalist account. But it is very unclear what a suppositional treatment of (13) might look like, which defenders of suppositional theories are normally quite willing to admit.

In view of these difficulties, maybe it is time to turn things around, and now it is for the one who takes (1) and (4) to be (almost) synonymous to bite the bullet and to dismiss the suppositional theory of acceptance for counterfactuals as being ill-founded despite its initial attractions. This is more or less what Bennett (2003) suggests, who defends the suppositional theory of indicative conditionals while maintaining at the same time a Lewisian theory of counterfactuals (for which (2) could count as something like a layman's rendering of the original Lewisian counterfactual, with 'necessarily' quantifying implicitly over all closest antecedent worlds). The downside of this proposal is that the suppositional acceptance of indicative conditionals in terms of conditional epistemic probability is very plausible indeed; and at the same time, indicative conditionals and subjunctive conditionals are treated very much alike in various circumstances, in particular, indicative and subjunctives about future events are very hard to distinguish on grounds of acceptance. Therefore, reserving a suppositional treatment *just* for indicatives would seem quite problematic. Furthermore, one can hardly deny that the epistemic operation of counterfactual supposition as described before does exist, and it would be quite surprising if that operation was completely independent of the epistemic acceptability of counterfactuals. In a nutshell: subjunctive conditionals seem to play *both a suppositional role*, as being given by their version of the Ramsey test, *and a descriptive role*, as being given by their truth conditions, and disregarding either of them would be a mistake.

So here is what we suggest as an attractive way out of this dilemma: *Counterfactuals carry both a suppositional degree of acceptability and a degree of belief-to-be-true, without any presumption that the two degrees have to be equal.* Thus, what actually went wrong in the derivation of our little triviality result above was the substitution of $Acc_{\mathfrak{Cr}_{\{w'\}}}(A \,\Box\!\!\rightarrow B)$ for $\mathfrak{Cr}_{\{w'\}}(A \,\Box\!\!\rightarrow B)$ when we moved from line 3 to line 4: we suggest to take this derivation not to show that either degrees of acceptability or degrees of belief have to be dismissed, but just that the degree of acceptability for a counterfactual should not be assumed to equal in value to the degree of belief for that counterfactual. We will now explain why this is not as desperate as it might seem. (If you don't think it is, even better!)

Whether one calls it a degree of acceptability or a degree of belief of a sentence, the idea is always to assign some number lying (not necessarily strictly) between 0 and 1 to that sentence. But of course this becomes a meaningful act only once the intended interpretation of that number has been specified. For a simple descriptive statement such as 'Ernest strikes a match (at time so and so)' this is more or less straightforward: The sentence expresses a semantic constraint on what our world is like; one can extensionalize this constraint by conceiving of it as a proposition or a set of possible worlds, that is, the set of worlds in which the sentence is true. That set is then assigned a number—a degree of belief—the pragmatic interpretation of which may be pinned down prototypically by turning to betting situations and to questions such as: Which betting quotients for bets on the truth of the sentence would the agent accept? Which bets on the truth of the sentence would the agent consider as fair? Accordingly, if the sentence is asserted by someone truthfully, then that person normally intends her audience to adopt her own high degree of belief in the truth of the asserted sentence: so additional to the semantic constraint, a pragmatic constraint on the epistemic system of the other agents gets expressed by an assertion of the sentence in normal conditions. However, this pragmatic constraint is derivative from the original semantic constraint, since the degree of belief for the sentence in question is determined on the basis of the set of worlds in which it is true. Call the semantic constraint the *semantic meaning* of 'Ernest strikes a match (at time so and so)' and the corresponding pragmatic constraint the *pragmatic meaning* of 'Ernest strikes a match (at time so and so)': [27] then the semantic meaning of that sentence is logically prior to its pragmatic meaning which gets determined only indirectly from it.

So much for 'Ernest strikes a match (at time so and so)'.[28] Now back to the topic: subjunctive conditionals. Once again we want to assign numbers between 0 and 1 to sentences, and we want these numbers to be meaningful pragmatically as being given prototypically by the function they have in betting situations. Let us consider a subjunctive conditional about a near future event: *If Ernest were to strike the match, it would light.* And say one is offered a bet on that conditional, that is, *someone bets that if Ernest were to strike the match, it would light.* Due to the peculiar features of conditional statements (or at least of those in the subjunctive mood), there are two permissible ways of interpreting that offer:

1. The person offers a conditional bet: *if Ernest were to strike the match, the person would bet that it would light.* If Ernest did not strike the match, the bet would be off; otherwise an unconditional bet on the match lighting would be initiated. Numerically, the bettor puts forward a conditional betting quotient which one may or may not accept.
2. The person offers an unconditional bet: *the person bets that the following proposition is true: the conditional chance of the match lighting given Ernest strikes it is*

[27] Paul Grice's seminal work is of course the primary source of the systematic study of pragmatic meaning or speaker meaning. However, in the following we are going to omit any discussion of implicatures. More recently, in dynamic semantics (Amsterdam style, usually stated as a nonprobabilistic theory), what we call pragmatic meaning here simply becomes identified with meaning itself, the meaning of a sentence being the change that incorporating the sentence in an epistemic system brings about (sometimes referred to as 'context change potential'); this guiding thought has also been applied to analyze the (pragmatic) meaning of counterfactuals: cf. Veltman (2005). Kratzer's premise semantics for counterfactuals in Kratzer (1981) is formally a variant of Veltman's, however it is presented as a truth-conditional semantics (indeed one that is equivalent to a variant of Lewis' semantics, as shown in Lewis, 1981; we thank Daniel Rothschild for help on this); and see van Benthem et al. (2009b) and Baltag & Smets (2008) for recent probabilified versions of such dynamic models. There is yet another type of dynamic analysis of counterfactuals, exemplified by the work of von Fintel (2001) and Gillies (2007), which is nonprobabilistic again and which also highlights the pragmatic contextual aspects of speaking in terms of counterfactuals. We will have to postpone any discussion of these dynamic approaches to another occasion.

[28] Linguistic expressions other than descriptive sentences behave quite differently: take 'Hooray!'. This expression does not express any semantic constraint at all in the sense of truth conditions or anything like that, however, uttering it does impose a pragmatic constraint on the audience, namely: *(if necessary) adjust your epistemic system in the way that you believe that I am happy.* This constraint is the pragmatic meaning of 'Hooray!'; and it is determined directly without recourse to any semantic meaning (cf. chapter 44 again in Bennett, 2003).

very high. Whether Ernest does strike the match or not is not the issue; in order to find out who would win the bet, one would need to investigate the physical condition of the match and its environment, Ernest's match-striking capabilities, and so on. Numerically, the bettor puts forward a standard unconditional betting quotient which one may or may not accept.

We suggest that the agent determines whether to accept the first bet and whether it is to be regarded as fair by comparing her *degree of acceptability* $Acc_{\mathfrak{Cr}}(A \mathbin{\Box\!\!\rightarrow} B)$ for that conditional $A \mathbin{\Box\!\!\rightarrow} B$ with the conditional betting quotient in question; indeed, in this "subjunctive bet on the future" type example, $Acc_{\mathfrak{Cr}}(A \mathbin{\Box\!\!\rightarrow} B)$ may be expected to coincide with the agent's conditional probability $\mathfrak{Cr}(B \mid A)$. But the agent determines whether to accept the second bet and whether it is to be regarded as fair by comparing her *degree of belief* $\mathfrak{Cr}(A \mathbin{\Box\!\!\rightarrow} B)$ in the truth of $A \mathbin{\Box\!\!\rightarrow} B$ with the unconditional betting quotient in question. Just as the two betting situations are different, the agent's degrees of acceptability and belief for one and the same conditional may differ, and indeed they have a different interpretation. Which degree ought to be used depends on the type of situation; presumably, in the betting example above, one would need to ask the bettor: *how exactly shall I understand this bet?* And we use the terms 'acceptability' and 'belief' to distinguish the two types of degrees. The triviality result above in fact proves that these degrees of acceptability and belief *could not* always line up perfectly, but that is alright as they measure two distinct phenomena: The Ramsey test governs the assignment of degrees of acceptability but not of degrees of belief to be true. Degrees of belief for counterfactuals are really degrees of belief for the propositions that are expressed by counterfactuals; the degree of acceptability of a counterfactual is not given by any assignment of a number to the proposition that is expressed by the counterfactual, rather on the propositional level the number comes attached to the pair of propositions expressed by the antecedent and the consequent of the counterfactual. A similar point could have been made in terms of bets on subjunctive conditionals about the past—*someone bets that if Ernest had struck the match, it would have lit*—however, bets on the past and especially on events which did not take place are not quite as common. But the structure of the argument would have been the same: in the conditional bet, if Ernest had not struck the match, the person bets it would have lit; in the unconditional bet, the person bets it is true that the conditional chance of the match lighting, say, 2 hours ago, if struck by Ernest back then, is very high. The latter bet could be on even if Ernest had not struck the match.[29]

Accordingly, although any counterfactual $A \mathbin{\Box\!\!\rightarrow} B$ has just one semantic meaning—the set of worlds in which the conditional chance of B given A is very high—it also has *two* pragmatic meanings:

1. In normal circumstances, asserting $A \mathbin{\Box\!\!\rightarrow} B$ expresses the pragmatic constraint that the audience ought to modify their epistemic systems in a way such that their

[29] A distinction between acceptance and belief is of course well-known from various other areas: Stalnaker (1984, p. 79f) makes it in his pragmatic theory of intentionality and conditionals, Bratman (1992) in his theory of intention and practical reasoning, van Fraassen (1980)in his theory of the constructive empiricist acceptance of scientific theories, Unwin (2007) in his epistemological answer to scepticism, and so forth. Although there are lots of significant differences among these accounts of acceptance versus belief, and also between them and the one that we employ in this section, what they all seem to have in common is that acceptance is some form of pragmatic commitment which does not aim at truth, in contrast with belief; see Unwin (2007, chap. 4), and Paglieri (2009) for more on this.

expected conditional chances for B given A become high. At the same time, on the side of the speaker, it is in normal circumstances a necessary condition for the assertability of $A \mathrel{\Box\!\!\rightarrow} B$ that the expected conditional chance of B given A is high for the one who is asserting $A \mathrel{\Box\!\!\rightarrow} B$.

2. In normal circumstances, asserting $A \mathrel{\Box\!\!\rightarrow} B$ expresses the pragmatic constraint that the audience ought to modify their epistemic systems in a way such that their degrees of belief in the conditional chance of B given A being very high become high. On the side of the speaker again, it is in normal circumstances a necessary condition for the assertability of $A \mathrel{\Box\!\!\rightarrow} B$ that the degree of belief in the conditional chance of B given A being very high is high for the one who is asserting $A \mathrel{\Box\!\!\rightarrow} B$.[30]

Call the former constraint the *direct pragmatic meaning* of a counterfactual, and the latter constraint its *indirect pragmatic meaning*. The indirect pragmatic meaning of $A \mathrel{\Box\!\!\rightarrow} B$ is determined indirectly again on the basis of the semantic meaning of $A \mathrel{\Box\!\!\rightarrow} B$, while the direct pragmatic meaning of $A \mathrel{\Box\!\!\rightarrow} B$ does not require any semantic meaning at all as input (other than the semantic meanings of A and B). With respect to indirect meaning, semantics is prior to pragmatics as the classic truth- conditional tradition has it, while for direct meaning semantics does not play a role at all (except for the semantic meanings of A and B), as the Pragmatist–Empiricist tradition has it. The suppositional account of counterfactuals tracks successfully the direct pragmatic meaning of counterfactuals, whereas the seemingly rivaling considerations on the pragmatics of conditional chance statements from above track their indirect pragmatic meaning. None yields a complete pragmatic account of counterfactuals without the other, none contradicts the other. In fact, they may even benefit from each other: For instance, once counterfactuals are seen to express propositions and to come with suppositional degrees of acceptability at the same time, it becomes possible to assign degrees of *acceptability* even to nested counterfactuals of the form

$$(A \mathrel{\Box\!\!\rightarrow} B) \mathrel{\Box\!\!\rightarrow} C$$

such as

(11) If it were the case that if the match were struck, it would light, then I could change that easily by dipping the match into water

from above. Since $(A \mathrel{\Box\!\!\rightarrow} B)$ expresses a proposition, which is its semantic meaning, the expected conditional chance of C given that proposition is well-defined, and hence the whole counterfactual $(A \mathrel{\Box\!\!\rightarrow} B) \mathrel{\Box\!\!\rightarrow} C$ may be assigned a degree of acceptability in the sense of the suppositional account of counterfactuals in spite of its antecedent being nested.

There is one remaining worry about this which needs to be addressed: Assume that $A \mathrel{\Box\!\!\rightarrow} B$ gets asserted in normal circumstances and not within the context of some other operator (such as the '... bets that ...' context from above): then the question arises *which* of the two pragmatic meanings—the direct one or the indirect one—is meant to be conveyed. One possible answer would be to make this dependent on the pragmatic context again. (Of course this would have to be worked out in detail.) However, and fortunately, it will usually not matter much, since both pragmatic constraints are about *high* degrees

30 In certain cases, some properties of the *manner* or *method* by which the receiver is meant to adapt their epistemic system might also be part of either of these pragmatic constraints, but we put this to one side here.

of acceptability or belief *and one can show that* $Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$ *is "high" if and only if* $\mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B)$ *is "high,"*[31] in a sense that can be made precise in either of two ways:

First, one can see immediately (without presupposing determinism in any way) that

$$\sum_{w} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) = 1 \text{ iff } \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) = 1\}) = 1$$

must hold. If we call $A \mathrel{\Box\!\!\rightarrow} B$ true at w again if and only if $\mathfrak{P}_w(B|A) = 1$, the proposition that is expressed by $A \mathrel{\Box\!\!\rightarrow} B$ is really $\{w : \mathfrak{P}_w(B|A) = 1\}$, and the credence of $A \mathrel{\Box\!\!\rightarrow} B$ is really the credence in that proposition. Hence we may reformulate this equivalence in terms of:

$$Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B) = 1 \text{ iff } \mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B) = 1.$$

Thus for 'high' in the sense of 'maximally high', the degree of acceptability of $A \mathrel{\Box\!\!\rightarrow} B$ is high if and only if the degree of belief in $A \mathrel{\Box\!\!\rightarrow} B$ is high.

Secondly, let us presuppose again some α in $[0, 1]$; in fact, in one part of Theorem 2.1 below, α will have to be greater than 0. We shall regard conditional chances $\geq 1 - \alpha$ as "very high," so typically think of α as some number less than $\frac{1}{2}$ (although we will not need this for the theorem). In the next section, we will say accordingly that $A \mathrel{\Box\!\!\rightarrow} B$ is approximately true to degree $1 - \alpha$ in a world w if and only if $\mathfrak{P}_w(B|A) \geq 1 - \alpha$. While for $\alpha > 0$ this notion of approximate truth will not coincide with what we will call truth *simpliciter*—that is, the case where $\mathfrak{P}_w(B|A) = 1$—as far as the differences between the two kinds of pragmatics meanings above are concerned, and whether or not they will show up when $A \mathrel{\Box\!\!\rightarrow} B$ is asserted, it should not matter much if we turn to the approximate version of "very high" conditional chances or to the "maximally high" one. For the theorem, we keep an atomic algebra over the given set W of worlds fixed, and then in the theorem we quantify simultaneously over credence functions $\mathfrak{Cr}$ (on that algebra) and over different possible assignments of conditional chance functions $\mathfrak{P}_w$ (over the same algebra) to worlds w in W:

THEOREM 2.1

- *Let* $0 < \alpha \leq 1$. *For every* $\epsilon > 0$ *there is a* $\delta > 0$, *such that for all epistemic probability measures* $\mathfrak{Cr}$, *and for all families* $(\mathfrak{P}_w)_{w \in W}$ *of world-relative conditional chance functions:*

  If $\sum_w \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) \geq 1 - \delta$, *then* $\mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1 - \alpha\}) \geq 1 - \epsilon$.

- *Let* $0 \leq \alpha \leq 1$. *For every* $\epsilon > 0$ *there is a* $\delta > 0$, *such that for all epistemic probability measures* $\mathfrak{Cr}$, *and for all families* $(\mathfrak{P}_w)_{w \in W}$ *of world-relative conditional chance functions:*

  If $\mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1 - \alpha\}) \geq 1 - \delta$, *then* $\sum_w \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) \geq (1 - \epsilon)(1 - \alpha)$.

For now, take the proposition that is expressed by $A \mathrel{\Box\!\!\rightarrow} B$ to be the set $\{w : \mathfrak{P}_w(B|A) \geq 1 - \alpha\}$ of worlds in which $A \mathrel{\Box\!\!\rightarrow} B$ is approximately true to degree $1 - \alpha$, so that the credence of $A \mathrel{\Box\!\!\rightarrow} B$ is really the credence in that proposition. Then what the theorem says is: if $Acc_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$ tends to 1 then $\mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B)$ tends to 1, and if $\mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B)$ tends to 1

[31] See Kvart (unpublished) for the suggestion to determine the assertability or acceptability of counterfactuals in terms of $\mathfrak{Cr}(A \mathrel{\Box\!\!\rightarrow} B)$ being high.

then $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B)$ tends to $1-\alpha$. 'Tending' here means: if one alters credence functions and world-relative conditional chance functions in the required way. In fact, the proof of Theorem 2.1 (which is included in the appendix) shows even more: One may choose $\delta = \epsilon \cdot \alpha$ in the first case and $\delta = \epsilon$ in the second one. Hence, if $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) \geq 1-\epsilon \cdot \alpha$ then $\mathfrak{Cr}(A \,\Box\!\!\rightarrow B) \geq 1-\epsilon$, and if $\mathfrak{Cr}(A \,\Box\!\!\rightarrow B) \geq 1-\epsilon$ then $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) \geq (1-\epsilon)(1-\alpha)$. For instance, for $\alpha = \epsilon = 0.05$, if $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) \geq 0.9975$ then $\mathfrak{Cr}(A \,\Box\!\!\rightarrow B) \geq 0.95$, and if $\mathfrak{Cr}(A \,\Box\!\!\rightarrow B) \geq 0.95$ then $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B) \geq 0.9025$. Therefore, even when 'high' does not quite mean 'maximally high', the degree of acceptability of $A \,\Box\!\!\rightarrow B$ turns out to be high if and only if the degree of belief in $A \,\Box\!\!\rightarrow B$ is high.[32]

What this means is that as far as the plain assertion of a counterfactual $A \,\Box\!\!\rightarrow B$ is concerned, whether it aims to bring about in the recipient the "highness" of $Acc_{\mathfrak{Cr}}(A \,\Box\!\!\rightarrow B)$ or rather the "highness" of $\mathfrak{Cr}(A \,\Box\!\!\rightarrow B)$ should not matter much, as long as "highness" is taken restrictively enough, and the same holds for the twofold necessary conditions on the assertability of counterfactuals. While generally the degree of acceptability for a counterfactual may differ significantly in value from the degree of belief for that counterfactual, differences like that are washed out more or less when one is stepping down from the quantitative scale of degrees of acceptability or belief to the qualitative one of assertion or assertability *simpliciter*. Hence, when a counterfactual is asserted (in normal circumstances), the two pragmatic constraints from above which derive from our two different kinds of acceptance/belief for counterfactuals amount to more or less the same constraint. Indeed, on the side of the recipient, if one were in complete and conscious control of one's epistemic system, and if one were to satisfy the requirement to change one's system in the way that the expected conditional chance of B given A becomes high, then it is very hard to see how this should be done plausibly and feasibly other than by determining the epistemically possible worlds in which the conditional chance of B given A is high and by subsequently raising their epistemic probabilities; that is, by assigning a high credence to the proposition that the conditional chance of B given A is very high. One might summarize this result in the way that while the direct and the indirect pragmatic meaning of counterfactuals differ intensionally, they coincide—more or less—extensionally.[33]

---

32 A similar remark can be made about the crucial step in the derivation of our little triviality result from above: there we had that

$$\mathfrak{Cr}((A \,\Box\!\!\rightarrow B) \wedge \{w'\}) = \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{Cr}_{\{w'\}}(A \,\Box\!\!\rightarrow B)$$

on the one hand, and

$$\mathfrak{Cr}(\{w'\}) \cdot Acc_{\mathfrak{Cr}_{\{w'\}}}(A \,\Box\!\!\rightarrow B) = \mathfrak{Cr}(\{w'\}) \cdot \mathfrak{P}_{w'}(B|A)$$

on the other, and by identifying degrees of belief and degrees of acceptability in value, it followed that indeterminism got excluded. Now we would not make this identification anymore, but what we would still be allowed to say is: $\mathfrak{Cr}(\{w'\}) \cdot \mathfrak{P}_{w'}(B|A)$ is "very close" to $\mathfrak{Cr}(\{w'\})$ if and only if $\mathfrak{Cr}((A \,\Box\!\!\rightarrow B) \wedge \{w'\})$ is "maximally close" to, that is, identical to, $\mathfrak{Cr}(\{w'\})$, where we take the proposition that is expressed by $A \,\Box\!\!\rightarrow B$ to be the set of worlds again in which $A \,\Box\!\!\rightarrow B$ is approximately true to degree $1-\alpha$.

33 For the same reason, if the pragmatic logic of simple counterfactuals of the form $A \,\Box\!\!\rightarrow B$, with neither A nor B including $\Box\!\!\rightarrow$ itself, is determined from their Skyrmsian assertability conditions in a way similar to how Adams did this for his logic for indicatives on the basis of his assertability conditions, then one can show that the resulting system of logic is but the corresponding flat fragment of the logic for $\Box\!\!\rightarrow$ that we will determine in the next section from the probabilistic truth conditions for counterfactuals. In fact, this flat fragment is simply Adams' logic for conditionals again. (Formally, this follows from the observation that an expected conditional chance is nothing but a convex combination of conditional probabilities.)

We resist the temptation to go full circle now and to return to the suppositional theory of *indicative* conditionals in order to see whether similar moves might be made also for them. Let it suffice to say that the argument could not be carried over without substantial adaptations: while both the suppositional pragmatics of indicative conditionals on the one hand, and the suppositional pragmatics as well as our truth-conditional semantics for subjunctive conditionals on the other, are all based on conditional probability, it is epistemic conditional probability in the case of indicatives and ontic conditional probability in the case of subjunctives. The latter is world relative and hence subserves the assignment of truth values relative to worlds; the former is not, except if one considers ways the agent's epistemic system and pragmatic context might be like and if one relativizes conditional epistemic probability measures to worlds that are centered on these additional parameters, such that indicative conditionals would be true or false at worlds relative to states of the agent's epistemic system and to communicative contexts.

Instead, we will now turn to the precise details of the semantics of counterfactuals that emerged from this section, which, as we saw, is a truth-conditional semantics in the traditional sense based on world-relative conditional chance.

§3. A new semantics for counterfactuals: the Popper function semantics. The basic idea of the semantics for counterfactuals that derived from the last section was very simple. Here is the strategy again: We let the truth of, for example,

If the match were struck, it would light

consist in its corresponding conditional probability, that is, the probability of the match lighting if struck, being very high or close to 1, where we are going to interpret this conditional probability as conditional chance again later. The semantics will therefore involve reference to, and quantification over, conditional probability functions.[34]

Usually, conditional probabilities $\mathfrak{P}(Y|X)$ are introduced on top of the absolute probabilities $P(Y \cap X)$, $P(X)$ by means of the Ratio formula:

$$\mathfrak{P}(Y|X) = \frac{P(Y \cap X)}{P(X)}.$$

Obviously, this is only possible if the absolute probability $P(X)$ is greater than zero. In contrast, there is also a well-developed and beautiful theory of primitive conditional probability measures or *Popper functions* (cf. Popper, 1968; Stalnaker, 1970; McGee, 1994; Roeper & Leblanc, 1999; Makinson, unpublished; and many more)[35] from which in turn absolute probabilities can be derived, and which allow for conditionalization on events or propositions with zero absolute probability. Clearly, conditionalization on absolute zero sets should not be disregarded by mere *fiat*. Indeed, as Hájek (2003) argues, conditional

[34] We shall speak of conditional probability *functions* and conditional probability *measures*—as well as absolute probability *functions* and absolute probability *measures*—interchangeably.

[35] Although these measures are usually called 'Popper functions', they have a history prior to, and independent of, Popper's work. As Horacio Arló-Costa reminded us in personal communication, the central Multiplication Axiom below is not originally due to Popper. It appears in Jeffreys before and in Keynes and apparently it was originally proposed by Johnson at Cambridge in a seminar that was attended by both Jeffreys and Keynes. These days Popper functions are very often put forward to be the "correct" formalizations of primitive conditional probability measures, but there is no consensus about this either: see Seidenfeld et al. (2001) and Easwaran (2005) for contrary views.

probability measures *ought* to be considered conceptually primitive rather than being reducible to absolute probability. As we will see later (cf. section 2.3 of part B), such conditional probability functions will allow us to interpret '$\mathfrak{P}(Y|X) = 1$' in two different ways: either *literally*, that is,

> the probability of Y given X is exactly 1 ("maximally high")

or *qualifiedly*, that is,

> the probability of Y given X is very high, that is, close to 1

where 'very high' and 'close' are *vague* terms. We are going to exploit the existence of these two possible readings later on.

But first things first. Let W be a (nonempty) set of possible worlds. Let $\mathfrak{A}$ be a Boolean field on W; we are going to interpret the members of $\mathfrak{A}$ as propositions, hence, propositions are taken to be sets of possible worlds again. Then we can define:[36]

DEFINITION 3.1 *$\mathfrak{P}$ is a conditional probability measure ("Popper function") on $\mathfrak{A}$ iff*

1. $\mathfrak{P} : \mathfrak{A} \times \mathfrak{A} \to [0, 1]$.
2. $\mathfrak{P}(X|X) = 1$.
3. *If* $\mathfrak{P}(W \setminus X \mid X) \neq 1$ *then* $\mathfrak{P}(.|X)$ *is a (finitely additive)*[37] *probability measure.*[38]
4. *Multiplication Axiom:* $\mathfrak{P}(X \cap Y|Z) = \mathfrak{P}(X|Z)\,\mathfrak{P}(Y|X \cap Z)$.
5. *If* $\mathfrak{P}(X|Y) = \mathfrak{P}(Y|X) = 1$, *then for all* $Z \in \mathfrak{A}$*:* $\mathfrak{P}(Z|X) = \mathfrak{P}(Z|Y)$.

Absolute probabilities may be defined on the basis of conditional probabilities by means of conditionalization on the uniquely determined tautological proposition, that is:

$$P(X) = \mathfrak{P}(X|W).$$

From the Multiplication Axiom 4 above it follows that if the derived absolute probability $\mathfrak{P}(X|W)$ is greater than zero, then

$$\frac{\mathfrak{P}(X \cap Y|W)}{\mathfrak{P}(X|W)} = \mathfrak{P}(Y|X \cap W) = \mathfrak{P}(Y|X)$$

which is nothing but a restatement of the Ratio Formula. However, even in the case $\mathfrak{P}(X|W) = 0$ the conditional probability $\mathfrak{P}(Y|X)$ is well-defined and not bound to be

---

36 Popper functions are sometimes introduced in the way that their second arguments are taken from a more restrictive class of propositions than their first arguments (as, e.g., in van Fraassen, 1976), but we will use the more straightforward definition stated in the main text.

37 We do not require countable additivity or σ-additivity, since it is questionable whether measures of chance have this property; we will return to this point briefly in section 2.1 of part B when we deal with representation theorems for Popper functions.

38 $\setminus$ is the set-theoretic relative complement operation; $W \setminus X$ is thus the set of worlds w in W which are not in X. We demand here that if $\mathfrak{P}(W \setminus X \mid X) \neq 1$—in words: X is *normal*—then for all Y, Z in $\mathfrak{A}$: if $Y \cap Z = \varnothing$ then $\mathfrak{P}(Y \cup Z|X) = \mathfrak{P}(Y|X) + \mathfrak{P}(Z|X)$. It is easy to prove from the axioms that X is *abnormal*, that is, $\mathfrak{P}(W \setminus X \mid X) = 1$, if and only if for all $Y \in \mathfrak{A}$: $\mathfrak{P}(Y|X) = 1$ (see van Fraassen, 1976, p. 419). Furthermore, X is abnormal if and only if for all normal $A \in \mathfrak{A}$: $\mathfrak{P}(X|A) = 1$ (see van Fraassen, 1976, p. 420). Abnormality is but the probabilistic counterpart of impossibility: conditionalization on an abnormal proposition yields 1 trivially, while conditionalizing an abnormal proposition on a normal one yields 0 trivially. Formally, only one set is *required* to be abnormal by the axioms (indeed, by the Multiplication Axiom 4): the empty proposition $\varnothing$.

trivial (such as being equal to 1 or 0). Hence, as promised, Popper functions allow for conditionalization on absolute zero sets.

In fact, within the present context it is not merely optional to take this special case into account as well: Let $\mathfrak{P}$ be interpreted as a conditional chance measure. If X is about an event in the past which did not take place, then the absolute chance of X *now* should be 0, but we would not want to say that the conditional chance $\mathfrak{P}(Y|X)$ *now* should therefore be undefined or trivially 1 or 0: Even if Ernest did not strike the match yesterday—when the match was dry, there was enough oxygen around, Ernest's hand was feeling a bit sore from tennis which would not have allowed him to strike the match as quickly as he would have normally, and so forth—the conditional chance *now* of it lighting up yesterday given Ernest had struck it back then should still be reasonably high, though maybe it would not be close to 1 because of Ernest's sore hand. Accordingly, even though I did not play darts yesterday, the conditional chance *now* of me having won against my brother yesterday given we had played is nevertheless about 0.5, since we are equally talented (or untalented) players, we were both in good shape yesterday, none of us had forgotten their glasses, and so on. There is no reason to believe that such conditional chances would not be just as objective and world dependent as the conditional chance *now* of the match lighting given Ernest were to strike it in 2 hours of time or the conditional chance *now* of me winning in the darts game against my brother if we were to play tomorrow. So something like a Popper function is *needed* for our purposes, since conditional chances on propositions with absolute chance 0 should not be trivialized in any way.[39] (We will have to say much more about the issue of time-relativity of chances in Section 4.3.)

Since, as we claimed we would argue in part B, '$\mathfrak{P}(Y|X) = 1$' may be taken to express that the probability of Y given X is close to 1, that is, either precisely 1 or merely close to 1 (without being precisely 1), '$\mathfrak{P}(Y|W) = 1$' may be taken to say that the absolute probability of Y is close to 1, and hence $\mathfrak{P}(Y|W) = 0$' may be interpreted as saying that the absolute probability of Y is close to 0. In this sense, Popper functions allow for conditionalization on sets with an absolute probability identical to, or merely close to, 0. Note that if the absolute probability $\mathfrak{P}(X|W)$ of X is (close to) zero and X is normal in the sense of $\mathfrak{P}(W \setminus X \mid X) \neq 1$, then the "logic" of Popper functions does not impose any constraints on how the absolute probability $\mathfrak{P}(Y|W)$ of Y relates to the conditional probability $\mathfrak{P}(Y|X)$ of Y given X; for instance, even if $\mathfrak{P}(Y|W) = 1$, this does not entail

39 Not taking this into account may easily lead to failed philosophical diagnoses. In Edgington (1997, p. 111), Edgington states that "The … condition on A, that it does not, now, have zero chance of being true, is necessary for the existence of an objective chance of B given A.". As we have just pointed out, this is wrong; and probably it was triggered by presupposing the classic Ratio formula to be constitutive of conditional probability. If Edgington had taken up the Popper function view on conditional chance, then she could not have concluded in her paper that the well-known Gibbardian "non-objectivity" phenomenon would apply to subjunctive conditionals about the past even though for subjunctive conditionals about the future it would not hold (or it only rarely would). For the argument that she puts forward for this bifurcation is that chances which are conditional on false contingent statements about the past lack any objective ground (and nontrivial value), while chances conditional on false contingent statements about the future usually do not. However, one should not constrain conditional chances in that manner, and one need not do so either. Consequently, the Gibbard problem fails to apply to counterfactuals about the past just as much as it fails for counterfactuals about the future. As far as chances conditional on physically or metaphysically impossible antecedents are concerned—which are not the cases that figure in Edgington's subjunctive translations of Gibbard's original indicative examples—see our Footnotes 38 and 53 in this paper, and also footnote 15 of part B. For a discussion of the much more prominent Gibbard cases for indicative conditionals, see Bennett (2003, sec. 34).

that $\mathfrak{P}(Y|X) = 1$ without any further probabilistic assumptions on X and Y. (However, if both $\mathfrak{P}(Y|W) = 1$ and $\mathfrak{P}(X|W) > 0$ are assumed, then the axioms do entail that $\mathfrak{P}(Y|X) = 1$.)

Given this, we can formulate our new probabilistic semantics for counterfactuals—which we call the *Popper function semantics*—as follows:

DEFINITION 3.2 $\langle W, \mathfrak{A}, (\mathfrak{P}_w)_{w \in W}, [\![.]\!]\rangle$ *is a Popper function model iff*

- *W is a nonempty set of possible worlds.*
- $\mathfrak{A} = \{[\![A]\!] \mid A \in \mathcal{L}\}$,
  where $\mathcal{L}$ is the unrestricted language of conditional logic[40] *with standard propositional connectives and with the additional conditional sign $\Box\!\!\rightarrow$.*[41]
  Read: $[\![A]\!]$ is the proposition expressed by A.
- *For every $w \in W$: $\mathfrak{P}_w$ is a Popper function on $\mathfrak{A}$.*
- *$[\![.]\!] : \mathcal{L} \to \wp(W)$ is subject to the following semantic rules:*
  - *Standard semantic rules for standard propositional connectives.*[42]
  - *The probabilistic truth condition for subjunctive conditionals:*

$$w \in [\![A \Box\!\!\rightarrow B]\!] \text{ iff } \mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = 1.$$

Read: A is true in w iff $w \in [\![A]\!]$.[43]

---

[40] So: If A and B are formulas, then $A \Box\!\!\rightarrow B$ is a formula, too, without any constraints whatsoever on A or B.

[41] We will not have anything to say about the interaction of $\Box\!\!\rightarrow$ with quantifiers in this paper.

[42] So, for instance: $w \in [\![\neg A]\!]$ iff $w \notin [\![A]\!]$; and so forth.

[43] By the defining clauses of a Popper function model, the algebra $\mathfrak{A}$ is only given relative to the semantic value mapping $[\![.]\!]$, the mapping $[\![.]\!]$ is only given relative to the Popper functions $\mathfrak{P}_w$, and the functions $\mathfrak{P}_w$ are only given relative to the algebra $\mathfrak{A}$ again. If this were a chain of definitions, then this would obviously be problematic. However, the definition above is really only a definition of the predicate 'Popper function model' which imposes simultaneous constraints on the components of Popper function models; so no circularity is involved. Indeed, there are various ways of seeing that this is unproblematic. Here is one: Assume that Popper function models had been explained as above except that every Popper function in every world had been defined on the full powerset algebra over the given set W of worlds; the values that the world-relative Popper functions would take at worlds could have been chosen arbitrarily as long as the axioms for Popper functions would have been satisfied. The $[\![.]\!]$ function would assign propositions to formulas in $\mathcal{L}$ in the standard recursive manner, just as this is the case in standard possible world semantics; if one preferred, one could also have fixed first a truth value assignment for all propositional letters at all worlds in W and then extended that assignment by the semantic rules (including the one for the new $\Box\!\!\rightarrow$ sign) to $[\![.]\!]$ on the whole of $\mathcal{L}$. Now let $\mathfrak{A}$ be set of sets of worlds that are generated from $\mathcal{L}$ by $[\![.]\!]$: that is, let $\mathfrak{A}$ be the set of all sets of the form $[\![A]\!]$ where A is in $\mathcal{L}$. This is a (countably infinite) Boolean field of sets. Finally, restrict the Popper functions at all worlds to just that subalgebra $\mathfrak{A}$ of the original powerset algebra. These restricted world-relative Popper functions, together with $\mathfrak{A}$ and $[\![.]\!]$, yield a Popper function model in the sense of Definition 3.2. So one can always generate our Popper function models from ones for which the original circularity worry does not even arise by looking only at the propositions that are actually assigned to formulas. The advantages of our definition of 'Popper function model' are that one can always interpret the arguments of the Popper functions in these models in terms of the given language, and it is also possible to apply the representation theorems of section 2 in part B to the Popper functions in these models (the representation theorems will presuppose countable algebras). We are grateful to David Makinson for a discussion on this.

According to this semantics, every possible world has its conditional probability measure, just as in modal semantics every possible world has its set of accessible worlds; conditional probability measures thus come as a package with possible worlds, as it were, though it is possible that different possible worlds share one and the same conditional probability measure. It is important to note that possible worlds according to the semantics above are worlds in the modern sense of modal semantics: that is, possible worlds are but points in a space; they need not have any internal structure such as traditional Carnapian state descriptions do. In particular, W above may include two numerically distinct worlds which are indistinguishable in terms of the evaluations of formulas relative to them; a fortiori, two numerically distinct worlds may well assign the same truth values to all formulas without $\Box\!\!\rightarrow$. And two distinct worlds w and w' may assign the same truth values to all formulas without $\Box\!\!\rightarrow$ while at the same time $\mathfrak{P}_w$ differs from $\mathfrak{P}_{w'}$: thus the truth values of counterfactuals are not assumed to supervene on the truth values of the statements that do not include $\Box\!\!\rightarrow$.[44] One could of course impose this kind of supervenience as an additional requirement for particular purposes, but it is not taken to be part of the logic of counterfactuals. Similarly, one could in principle require all worlds to have one and the same conditional probability measure, but once again we do not demand this anywhere in the semantics itself, and the demand would also trivialize the evaluation of nestings of counterfactuals to a large extent since then the proposition expressed by whatever counterfactual would either be W or the empty set.

Note that in the semantics, every world-relative conditional probability measure is defined on one and the same set of propositions which can be expressed by statements in our given language $\mathcal{L}$. A counterfactual is true in a world if and only if its corresponding conditional probability is 1, as being determined by the Popper function of that world. Moreover, as we promised to explain in section 2.3 of part B, we will also be allowed to read our semantics clause for counterfactuals in the way: a counterfactual is true in a world if and only if its corresponding conditional probability is close to 1, where 'close to' is a vague term. Since Popper functions are world relative and since counterfactuals are evaluated at worlds, counterfactuals express propositions in the standard sense of possible worlds semantics, and nesting them and applying propositional operators to them is straightforward. This is the semantics *in nuce*. We will discuss its plausibility as a semantics for counterfactuals in natural language in the remainder of this article.

As always in a truth-conditional semantics, we can define a formula A in $\mathcal{L}$ to be logically true in the semantics if and only if A is true in every world in every model, that is in our case, in every Popper function model.[45] The resulting system of conditional logic is determined by the following theorem (the proof of which is to be found in the appendix):

THEOREM 3.3 *The system V of conditional logic (cf. Lewis, 1973b) is sound and complete with respect to the Popper function semantics for subjunctive conditionals:*

- Rules of V:
    1. Modus Ponens (for $\supset$)

[44] We thank Franz Huber for a helpful discussion on this.

[45] Accordingly, the logical consequence or entailment relation that holds between finite sets of formulas and single formulas can be defined in the standard way, as can be the concept of logical validity of arguments with finitely many premises and a single conclusion.

2. Deduction within subjunctive conditionals: for any $n \geq 1$

$$\frac{\vdash (B_1 \wedge \ldots \wedge B_n) \supset C}{\vdash ((A \mathbin{\Box\!\!\rightarrow} B_1) \wedge \ldots \wedge (A \mathbin{\Box\!\!\rightarrow} B_n)) \supset (A \mathbin{\Box\!\!\rightarrow} C)}$$

3. Interchange of logical equivalents

- Axioms of V:

  1. Truth-functional tautologies
  2. $A \mathbin{\Box\!\!\rightarrow} A$
  3. $(\neg A \mathbin{\Box\!\!\rightarrow} A) \supset (B \mathbin{\Box\!\!\rightarrow} A)$
  4. $(A \mathbin{\Box\!\!\rightarrow} \neg B) \vee (((A \wedge B) \mathbin{\Box\!\!\rightarrow} C) \leftrightarrow (A \mathbin{\Box\!\!\rightarrow} (B \supset C)))$

Accordingly, neither Strengthening of the Antecedent (Monotonicity) nor Transitivity nor Contraposition are logically valid rules. V is in fact slightly weaker than Lewis' preferred logic VC for counterfactuals; we will have much more to say about this in section 1 of part B.

Although rules such as Strengthening of the Antecedent are neither valid by the lights of our semantics nor by Lewis', the reasons for it being invalid are of course not quite the same, due to the differences between the two semantics: in Lewis' case,

> If you had walked on the ice, it would have broken

is perhaps true but

> If you had walked on the ice while leaning heavily on the extended arm of someone standing on the shore, the ice would have broken

is not, since worlds in which you were walking on ice without leaning on anyone like that are much closer to the actual world than worlds in which you were walking on ice while leaning on someone as stated.[46] In the probabilistic semantics, maybe, the former counterfactual is true because the probability of the ice-breaking conditional on you walking on it was close to 1, though the latter counterfactual might still be false if only walking on the ice without the leaning was so much more likely than walking on the ice with the leaning. But this already shows the differences between Lewis' semantics and the probabilistic semantics quite clearly: after all, degrees of similarity are not supposed to correspond to magnitudes of probability in any straightforward manner. At the same time, in many application cases, they should not be too far apart either: for instance, if the absolute chance of you walking on the ice in 5 minutes while leaning on someone as sketched were not close to 0, then it would very questionable if 'If you were to walk on the ice in five minutes, it would break' were actually true even from the viewpoint of Lewis' semantics. We will deal both with such coincidences and divergences between a similarity/closeness based semantics and our probabilistic semantics throughout this paper.

Theorem 3.3 is an extension of results by Hawthorne (1996) and Arló-Costa & Parikh (2005), who pursue a similar kind of probabilistic semantics for nonmonotonic consequence relations.[47] Since such consequence relations are expressed only metalinguistically, their semantics corresponds to a restriction of the Popper function semantics to the flat

---

[46] We take this example from Bennett's (2003) discussion on p. 160.

[47] Some of their formal results can be found already in McGee (1994), though expressed in terms of (nonnested) conditionals rather than in terms of nonmonotonic consequence relations.

fragment of the language of conditional logic in which neither nestings of conditionals nor applications of propositional operators to conditionals are permitted. By contrast, Theorem 3.3 applies to the full language of conditional logic.

Now we add a corresponding semantics for the *approximate truth* of counterfactuals. Let $0 \leq \alpha < \frac{1}{2}$:

DEFINITION 3.4 $\langle W, \mathfrak{A}, (\mathfrak{P}_w)_{w \in W}, [\![.]\!]_\alpha \rangle$ *is an* α*-approximate Popper function model iff the same clauses as in Definition 3.2 hold, except that '*$[\![.]\!]_\alpha$*' replaces* $[\![.]\!]$ *everywhere, and the semantic rule for counterfactuals is changed into:*

$$w \in [\![A \;\Box\!\!\rightarrow B]\!]_\alpha \text{ iff } \mathfrak{P}_w([\![B]\!]_\alpha \mid [\![A]\!]_\alpha) \geq 1 - \alpha.$$

Read: A is approximately true to degree $1 - \alpha$ *in* w *iff* $w \in [\![A]\!]_\alpha$.

Just as in the truth semantics from before, the approximate truth of a counterfactual at a world is tied to its corresponding conditional chance being very high. However, now 'very high' is interpreted neither as 'equal to 1', nor as 'close to 1' with 'close to' being vague, but rather a precise real-valued threshold $1 - \alpha$ is introduced, such that $A \;\Box\!\!\rightarrow B$ is approximately true in w if and only if $\mathfrak{P}_w([\![B]\!]_\alpha \mid [\![A]\!]_\alpha)$ exceeds or is equal to $1 - \alpha$. More precisely, depending on the α chosen, we speak of the approximate truth of $A \;\Box\!\!\rightarrow B$ *to degree* $1 - \alpha$, where one should think of α as being small. We demand that $\alpha < \frac{1}{2}$ as this ought to be the minimal requirement on 'small'.

There are three ideas that motivate this additional notion (or family of notions) of approximate truth: First of all, it corresponds to an alternative way in which the truth of a counterfactual may be understood in terms of its corresponding world-relative conditional probability being very high. Secondly, we have seen this notion of approximate truth, and the parameter α on which it is based, successfully at work in Theorem 2.1 of the last section. Thirdly, it is a well-known problem in the semantics of scientific languages that scientific hypotheses are rarely true in the strict Tarskian sense of the word. More regularly, empirical hypotheses or theories entail numerical predictions which coincide with the empirical data only approximately, which is why these hypotheses or theories are called approximately true rather than true *simpliciter* (see Resnik, 1992 for an overview). In general, the precise explication of 'approximate truth' is quite elusive, however for plain atomic statements about numerical values the intended interpretation of 'approximately true' is quite straightforward: 'Peter is 6 feet tall' is true if and only if Peter is 6 feet tall. Accordingly, 'Peter is 6 feet tall' is approximately true if and only if Peter is 6 feet tall *or not far from* 6 *feet*, where the metric distance between Peter's actual height and the "predicted" 6 feet covary with the imagined "distance from the truth," but where the truthmaker of the sentence is the same for both truth and approximate truth, that is, Peter's numerically given height. What the approximate truth part of our semantics above does is merely to translate this into the context of counterfactuals: the truthmaker of a counterfactual is its corresponding conditional probability[48]—which is measured numerically again—both in the truth and the approximate truth portion of our semantics; and approximate truth consists in a slight divergence of that numerical value from the value that would yield truth *simpliciter* again, that is, in this case: 1. Indeed, approximate truth to the maximal degree 1 *is* just truth *simpliciter* again. We will thus be able to say that a

[48] Kvart (unpublished) also regards "counterfactual probabilities" as the truthmakers of counterfactuals, and he suggests truth conditions for counterfactuals which are very close to those of our approximate truth semantics.

particular counterfactual is approximately true (relative to some fixed α) in the same sense in which atomic scientific statements about numerical values are said to be approximately true (relative to some fixed α). The threshold value in question could be held fixed through different pragmatic contexts of utterance and theorizing, but it could also be determined contextually by the current demands of the agents involved, though we will not have to say anything about this in further detail. Note that while it is was quite obvious how to take this step towards approximate truth in the present probabilistic, and hence quantitative, context, it would not be clear at all how a similar step could be taken in a qualitative semantics such as Lewis'.

A formula A in $\mathcal{L}$ is α-logically true in the semantics iff A is approximately true to degree $1 - \alpha$ in every world in every α-approximate Popper function model. Logically, the following extends results by Hawthorne (2007) and Hawthorne & Makinson (2007) to the full language of conditional logic again (as they consider nonmonotonic consequence relations again):

THEOREM 3.5 *For every $0 \leq \alpha < \frac{1}{2}$, the system Q^{*}[49] of conditional logic is sound with respect to the α-Popper function semantics for subjunctive conditionals.*

- Rules of Q*:

  1. Modus Ponens (for $\supset$) and Interchange of Logical Equivalents
  2. Right Weakening:
  $$\frac{\vdash B \supset C}{\vdash (A \mathrel{\Box\!\!\rightarrow} B) \supset (A \mathrel{\Box\!\!\rightarrow} C)}$$
  3. Very Cautious Monotonicity:
  $$\frac{\vdash A \mathrel{\Box\!\!\rightarrow} B \wedge C}{\vdash A \wedge B \mathrel{\Box\!\!\rightarrow} C}$$
  4. Weak Or:
  $$\frac{\vdash A \wedge B \mathrel{\Box\!\!\rightarrow} C,\ \ A \wedge \neg B \mathrel{\Box\!\!\rightarrow} C}{\vdash A \mathrel{\Box\!\!\rightarrow} C}$$
  5. Weak And:
  $$\frac{\vdash A \mathrel{\Box\!\!\rightarrow} B,\ \ A \wedge \neg C \mathrel{\Box\!\!\rightarrow} C}{\vdash A \mathrel{\Box\!\!\rightarrow} B \wedge C}$$

- Axioms of Q*:

  1. Truth-functional tautologies
  2. $A \mathrel{\Box\!\!\rightarrow} A$
  3. $(\neg A \mathrel{\Box\!\!\rightarrow} A) \supset (B \mathrel{\Box\!\!\rightarrow} A)$
  4. $(A \mathrel{\Box\!\!\rightarrow} C) \supset ((A \wedge B \mathrel{\Box\!\!\rightarrow} C) \vee (A \wedge \neg B \mathrel{\Box\!\!\rightarrow} C))$
  5. $((A \mathrel{\Box\!\!\rightarrow} \neg B) \wedge (A \mathrel{\Box\!\!\rightarrow} B)) \supset (A \mathrel{\Box\!\!\rightarrow} \neg A)$

Note that axiom scheme 5 is the only one that actually requires $\alpha < \frac{1}{2}$. More details on the formal properties of such probabilistic above-threshold logics can be found

[49] This name is not to be found in either Hawthorne (2007) or Hawthorne & Makinson (2007), but the present system is so closely related to their systems that we chose a name that is similar to some of theirs.

in Hawthorne (2007) and Hawthorne & Makinson (2007). Most notably, the And Rule (Conjunction Rule) or Agglomeration, that is,

$$A \mathrel{\Box\!\!\rightarrow} B, A \mathrel{\Box\!\!\rightarrow} C \vdash A \mathrel{\Box\!\!\rightarrow} B \wedge C$$

is valid according to our truth semantics from before but not according to the α-approximate truth semantics (for any $\alpha > 0$). Metaphorically speaking, one could view of the logic above as what remains of the system V if one "subtracts" Agglomeration from it. We do not know if the system above is also complete with respect to the new semantics in the sense that every formula which is α-logically true for every $0 \leq \alpha < \frac{1}{2}$ is provable in the system above, but the results contained in Paris & Simmonds (2009) suggest this to be unlikely.

Just as scientists do, we will focus mainly on *truth simpliciter* in the following, and we will retreat to *approximate truth* only when we must, in particular, when it becomes necessary for scientific reasons. But for now, if we take the truth part of our semantics—with its second probabilistic understanding in terms of the vague expression 'very high probability'—together with the approximate truth part of the semantics, we find that the truth (see (i) below) or approximate truth (see (ii) below) of

If the match were struck, it would light

at a world w allows for exceptions in either of the following two senses: (i) The conditional chance of the match lighting if struck is close to 1 at w with 'close to' being a vague term, which allows for the conditional chance at w of the match not lighting if struck to differ from zero (though it must be close to zero then). In section 2.2 of part B we will strengthen this case by showing that this may actually be understood as: independent of how close one gets to w, the truth of the counterfactual above allows for a positive conditional chance of the match not lighting if struck (though then this chance must tend to zero in the limit). (ii) The conditional chance of the match lighting if struck is greater than $1 - \alpha$ at w, with $0 \leq \alpha \leq \frac{1}{2}$, which allows for the conditional chance at w of the match not lighting if struck to exceed zero by any number between 0 and α (though it must not be α itself or greater).

§4. Interpreting the semantics.

4.1. The purpose of the semantics. Before we turn to the question of how to understand this semantics properly, and that is, how to understand its truth condition for counterfactuals properly—including, amongst others, the question of what kind of probability exactly it involves—an even more urgent prior question comes to mind: What exactly is the purpose of this semantics?

Here is the worry:[50] Its purpose *could not be* to give a conceptual analysis of the commonsensical $\Box\!\!\rightarrow$ in natural language—of what we have in mind when we think or utter it—since $\Box\!\!\rightarrow$ could not be *that* complex. Even though our semantics is quite simple as regards its underlying idea, maybe the notion of conditional probability is not. It might be unrealistic to assume that if one "zoomed" into the meaning of everyday counterfactual claims, one would necessarily find difficult concepts such as *conditional probability* or *being close to 1* somewhere within them; maybe one should not expect to find probabilistic concepts there at all, independent of their complexity. Furthermore, maybe competent

[50] We thank Tim Williamson for highlighting this as a problem to be addressed.

English speakers do in fact *not* allow for the existence of exceptional circumstances when they assert a statement of the form $A \mathrel{\Box\!\!\rightarrow} B$ to be true, even though the truth of $A \mathrel{\Box\!\!\rightarrow} B$ according to the probabilistic semantics does allow for exceptions. So what is the status of a semantics such as the above vis-à-vis the actual usage of counterfactuals in natural language?[51]

Indeed: We do not propose our semantics to yield anything like a traditional conceptual analysis of subjunctive conditionals. As a first approximation, the semantics rather aims to give the right truth conditions for $A \mathrel{\Box\!\!\rightarrow} B$. But this does not necessarily involve expressing all and only the "right" concepts that are underlying $\mathrel{\Box\!\!\rightarrow}$: In order to see this, take whatever a successful conceptual analysis of an ordinary $A \mathrel{\Box\!\!\rightarrow} B$ statement might look like[52] and add the sentence '$1 + 1 = 2$' to it conjunctively. $A \mathrel{\Box\!\!\rightarrow} B$ is then true according to this deviant analysis under the very same conditions under which is true in the original intended analysis, simply because '$1+1 = 2$' is true unconditionally. Clearly, it would be surprising if concepts for natural numbers and for natural number operations and relations played a role in what an English speaker has in mind when he or she asserts

> If the match were struck, it would light.

But this does not make it less true that also in the second analysis the conditions under which this counterfactual are true are stated correctly. So one can aim at the right truth conditions without aiming at the right conceptual analysis, which is what we do.

What is more, we even allow for deviations from the actual truth conditions of our everyday $\mathrel{\Box\!\!\rightarrow}$ if this leads to a better philosophical theory. Roughly, we submit to the old-fashioned ideal of Carnapian explication: a semantics for x obviously should not ignore paradigm case occurrences of x in actual bits of communication, but where philosophical theorizing needs to correct such occurrences in order to avoid philosophical problems or to make the semantics more continuous with science, this ought to be done without bad conscience. That is also the reason why we would be suspicious of any philosophical methodology that took natural language examples of x, counterexamples, counterexamples to these counterexamples, and so forth, too seriously. Obviously, having at hands good examples is nice, but if a philosophical theory of x is sufficiently elegant, precise, and coherent with established theories in science and philosophy, then this should trump the need to conform to *all* example cases of pretheoretical verdicts concerning x. (But of course such a theory should not contradict *all* such pretheoretical verdicts either.) In any case, we take the liberty to side with the Carnapian methodology even for $x = \mathrel{\Box\!\!\rightarrow}$.

In some cases, such deviations from pretheoretic everyday usage might even be necessary for more serious reasons. What if our commonsense $\mathrel{\Box\!\!\rightarrow}$ does not have a reference at all? It might aim at referring to some structure—say, as the standard semantics has it, one of Lewis' sphere systems or Stalnaker's selection functions—but by sheer misfortune there is simply no such structure to refer to. Or: What if our commonsense $\mathrel{\Box\!\!\rightarrow}$ does have a referent but our intuitive semantic theory of it, of which Lewis' semantics maybe is a systematic reconstruction, simply misdescribes what it refers to? It might be that while

---

51 If one *does* think that the concept of conditional probability is simple enough, that there are good reasons to believe that our ordinary $\mathrel{\Box\!\!\rightarrow}$ is probabilistic in nature, and that one typically understands counterfactuals with some *ceteris paribus* clauses attached to them, then our semantics should be attractive even if put forward as a conceptual analysis. If you think so, you might just as well jump to the next section directly.

52 If there is one—it might well be that our natural language $\mathrel{\Box\!\!\rightarrow}$ is regarded best as being conceptually primitive.

there is indeed some structure out there that our everyday $\Box\!\!\rightarrow$ hits upon semantically, what that structure is and how $\Box\!\!\rightarrow$ clings to it may be specified more accurately in terms of concepts which do not show up in the actual conceptual analysis of $\Box\!\!\rightarrow$ but which rather originate from independent scientific contexts. In either case, a thorough semantic revision of our previous understanding of $\Box\!\!\rightarrow$ would be imperative, or otherwise one would be guilty of hanging on to an empty or at least misleading semantic analysis.

This is thus what we hope our probabilistic semantics for counterfactuals will achieve: to state what the correct truth conditions for many occurrences of counterfactual statements in natural language ought to be like. We do not maintain that this probabilistic semantics gives the right analysis of the meaning that we normally attach to $\Box\!\!\rightarrow$ in everyday discourse. At best, we can hope that on the basis of the quantitative concept of conditional probability that our theory employs we will nevertheless be able to approximate a qualitative notion of conditional probability which is not too remote from the qualitative, though nonprobabilistic, conception that is likely to be underlying the competence theory for $\Box\!\!\rightarrow$ in natural language. But more importantly we hope that the rest of this section, in conjunction with sections 1 and 2 of part B, will show that this semantics has the virtues of a good philosophical theory which aims to find some middle ground between the manifest and the scientific image, between the metaphysical objects and concepts that are presupposed by Lewis' semantics and the objects and concepts that are presupposed by modern science. And we take section 3 in part B to demonstrate that our new probabilistic semantics for counterfactuals overcomes at least some of the problems that science, or philosophical insight inspired by science, poses to Lewis' semantics.

While we hope the semantics will ultimately turn out to have some nice applications to some more traditional metaphysical questions, we are far from claiming that all modal notions in metaphysics will eventually have to give way to probabilistic versions thereof. In particular, we do not think that our semantics has anything interesting to say about possibility, necessity, or strict implication.[53]

4.2. The intended interpretation of probability. Now back to the main semantic rule of our new semantics, that is:

$$w \in [\![A \Box\!\!\rightarrow B]\!] \text{ iff } \mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = 1.$$

We said that the function $\mathfrak{P}_w$ that is referred to by the right-hand side of this semantic clause satisfies the axioms of primitive conditional probability measures or Popper functions. But obviously this leaves such conditional probability measures heavily underdetermined. The obvious question to ask at this point is: What kind of probability do we have in mind here as far as *intended* Popper function models are concerned?

As discussed already in Section §2, since the conditionals that we are interested in are counterfactuals, which are at least *prima facie*—and in the typical case—sentences which describe the world and which are true or false, the probability measures in our semantics

[53] To be sure, if the underlying set W of possible worlds in our semantics is taken to be the set of metaphysically possible worlds, and if the only proposition X that is "abnormal" with respect to $\mathfrak{P}_@$, that is, for which $\mathfrak{P}_@(W \setminus X \mid X) = 1$ holds, is assumed to be the empty proposition, then $\Diamond A$ may in fact be expressed by means of $\neg(A \Box\!\!\rightarrow \bot)$ as far as the actual world @ is concerned. But that should not be understood as saying that possibility *reduces* to a probabilistic notion in any sense. Rather one would need to have an antecedent understanding of 'possible' and 'impossible' which one would then be able to import into the probabilistic theory; the latter would thus presuppose these notions, rather than explain them.

should be objective nonepistemic ones. When we say 'objective probabilities' here, we mean 'probabilities which are independent of the beliefs of persons or groups of persons'; when we say 'nonepistemic probabilities', we mean 'probabilities which are not measures of the degree of reasonableness of believing something'; see Achinstein (2001, chap. 5) for more on this terminology. Because the A and B in $A \mathrel{\Box\!\!\rightarrow} B$ are meant to be substitutable by complete sentences in the logical sense of the word, these objective nonepistemic conditional probability measures should yield conditional single case probabilities, in other words—according to the standard options for interpreting probability—*conditional single-case chances*.

There would be other options as well: Loewer (2007) presents a probabilistic account of subjunctive conditionals in terms of what he calls *statistical mechanical probabilities* which he also interprets as objective nonepistemic probabilities. The idea is that any given worldly macrostate underdetermines the underlying microstates of the world and hence only determines a probability measure over the microstate that are compatible with that macrostate (see also Albert, 2000). Loewer does not give the measures in question an epistemic interpretation in terms of our ignorance of microstates, but he rather interprets them as empirically supported law-governed objective probability measures of ontic origin. While we will stick to the standard interpretation of objective nonepistemic single-case probabilities in terms of single-case chances, will return to Loewer's paper in Section 4.4.[54]

For instance, in order to evaluate

> If the match (at location x) were struck (at time t), it would light

we suggest to turn to the

> conditional chance of the match (at location x) lighting given that it is struck (at time t),

[54] There are still other options. A different type of objective nonepistemic single-case probability would be *semantic probability*. Here is the idea: Let W be the set of conceptually or linguistically possible worlds; in traditional terminology, a sentence would be analytic if and only if it were true in all such possible worlds. But now assume that our conceptual apparatus or our language comes *itself* with a probability measure on the set of propositions on W, such that this measure generalizes the set of analytical statements to what might be called an "analytic probability measure." Accordingly, while the antecedent of

> If the match were struck, it would light

leaves open various ways in which the match could be struck—ways which are all consistent with the meaning of 'the match is struck'—competent English speakers automatically, and in virtue of their being speakers of English, might regard some ways of striking a match as more salient than others, some ways of striking a match as equally salient as others, some ways of physically interacting with a match as only partially striking a match, and so forth, where salience and partiality are coded in terms of probability. Of course, this kind of semantic probability is not on the map of standard interpretations of probability these days (except perhaps for probabilistic theories of vague concepts), but that does not mean that it could not be; moreover, the traditional interpretation of probability as *logical* is not too far from it. (And maybe Edgington's *verities* in Edgington, 1996, might be regarded as such *semantic probabilities*?) If such semantic probability measures could be made sense of, and if they were studied in more detail, then maybe they might even be found to be "mixed" into standard ontic probability measures or chances so that the assignment of objective nonepistemic single case probabilities would in fact be determined by wordly matters and semantic saliency simultaneously. In any case, for our current purposes it suffices to note that other intended interpretations of the conditional probability measures of our Popper function models might be available.

which at a possible world w would be determined by applying the Popper function $\mathfrak{P}_w(.|.)$ of that world to the relevant antecedent and consequent propositions. Thus, $\mathfrak{P}_w(.|.)$ should not be regarded subjective or epistemic.[55] We say 'should', as this is only the default assumption: if one believes to have a good theory of how to, say, analyze single case chances in terms of rational degrees of belief or the like, then please be our guest. The theory of single-case chances is a complicated matter and steeped in a history that in the current paper we want to avoid delving into for as much as we can. For instance, for Lewis, world-relative chances are ultimately determined by what the best overall theory of the actual world says they are like; for Skyrms, chances are resilient epistemic probabilities which are determined by conditionalizing an epistemic probability measure on the cells of a "natural" partition of W, where this partition in turn is determined on grounds of some robust scientific theory about the world; for others they are propensities of experimental setups to lead to particular results in specific circumstances; and there are many more interpretations. At this point we invite pretty much everyone to plug in his or her favorite theory of chance in order to support this semantics in their preferred way, but in the meantime we will continue to speak of single case chances as suggested. Indeed, we will speak of the chance of a proposition only relative to a possible world—along the lines of, for instance, Lewis and Skyrms on chance, who also relativize chances to worlds; accordingly, the chance of, for example, a particular match lighting if struck may vary from one world to the next, which is plausible enough in view of the fact that one and the same match might be more or less wet in different worlds, there might be different concentrations of oxygen around the match according to different worlds, and the like.[56]

In fact, the situation is similar with respect to single case chances in science: *Prima facie* it seems that quantum physics assigns objective nonepistemic probabilities to event tokens or to propositions describing single events, where the determination of these probabilities is a purely "worldly" matter. Of course, some physicists or philosophers of physics might argue that these probabilities are really other kinds of probability in disguise—maybe statistical probabilities of event types or rational credences of propositions or something else. But at least at first glance scientists seem to refer to the

[55] Hence, Joyce's concern (cf. Joyce, 1999, p. 202f) that "the sort of supposition represented by Rèyni-Popper measures is not in any sense subjunctive" does not affect the semantics above, since, in contrast with Joyce's presuppositions, our Popper functions $\mathfrak{P}_w(.|.)$ are neither taken to be subjective evidential conditional probability measures nor epistemic measures of subjunctive supposition based on probabilistic "imaging" (as defined in Lewis, 1976). However, there are interesting formal connections between conditionalization by means of Popper measures and variants of imaging: indeed one can show that Popper measures are special instances of Levi's and Gärdenfors' so-called "preservative imaging" functions; see Gärdenfors (1988, p. 117) for a definition. Furthermore, as mentioned in Section §2, Skyrms' update operation in terms of expected conditional chance, where conditional chances at worlds would be given by Popper functions again, is the generalization of imaging to the indeterministic case.

[56] Concerning the worlds themselves: In a purely semantic context such as the current one, the intended underlying set of worlds should be thought of as the set of physically or perhaps metaphysically possible worlds, while in an epistemic context—such as the one of Section §2—the intended set of worlds would be the set of epistemically possible worlds relative to a given agent and a point of time. In neither case we want to commit ourselves to any particular view of what a possible world is and which possible worlds there are; in particular, no special metaphysical view of worlds is needed. Indeed, a more scientific account of worlds as theoretical constructs, according to which the set of possible worlds is but the space of theory-relative complete parameter settings consistent with a scientific theory would do just as well.

chance that the atom (at location x) decays (in period $[t, t + \Delta t]$)

as if they referred to single case chances in the standard sense.[57]

Apart from general worries concerning such objective nonepistemic single case probabilities, there are more specific and well-known worries concerning *conditional* probabilities of that kind. One problem to be dealt with is Humphreys' (1985) according to which it is questionable whether conditional propensities can be inverted. For the moment, however, we simply assume that there is a plausible analysis of invertible conditional measures of chance—so that whenever '$\mathfrak{P}(E|F)$' is meaningful, '$\mathfrak{P}(F|E)$' is so, too—and that, accordingly, conditional single case chances are *not necessarily* tied to the antecedent of a subjunctive conditional describing the cause of a state or event that gets described by the consequent of that conditional—which is not to say, of course, that conditional chances would be independent of causal matters.

But overall the presumptions that we make on chance are fairly modest really: chances satisfy the formal axioms of Popper functions; chances are world-relative; chances are time-dependent (more about which in the next section), and in particular the chance *now* of a proposition about the past is simply the truth value of that proposition since there is no chance anymore of it having a different truth value, even though its chance *back then* of being true might have been strictly between 0 and 1; chances are related to epistemic probabilities as expressed by well-known reflection principles (which we will use in section 1.3 of part B); conditional chances are invertible in the sense explained before; causal independence entails chance independence (which will be important in Sections 4.3 and 4.7); and whatever philosophical theory of chance is used in order to support our Popper function semantics, it has to be a theory that is roughly in line with scientists' talk of chances within their scientific theories. That is what we presuppose, and there is actually a variety of quite different theories of chance from which all of these properties would be derivable. No particularly demanding metaphysical account of chance needs to be taken for granted by the intended interpretation of the semantics.

4.3. The time-relativity of chance and subjunctive conditionals. Even though we try not to enlarge upon the philosophy of chance beyond what we actually need to, there is one aspect of chance which we simply cannot ignore: its time-relativity. Standardly, people speak of chances as being indexed—whether implicitly or explicitly—by points of time; for instance, Lewis does so. The chance of Caesar winning at Alesia is 1 *now* but perhaps it was less than one at the beginning of the battle back in 52 BC, and it might have had yet another value a couple of months before the battle. Or using Lewis' (1980) well-known example: one chooses turns in a labyrinth by tossing a coin at each branching point; when entering the labyrinth there is a reasonable chance of reaching the center by noon, which then drops significantly when one strays into a region from which it is hard to reach the center, but then one turns lucky and gets closer to the center again, and after one's reaching the center just a little before noon the chance of one's reaching the center by noon is 1 and remains to be such.

This time-relativity feature of chance has been turned into an objection to truth-conditional semantics for counterfactuals that are based on chance such as the Popper function semantics of section §3: As Edgington (2008, p. 13), points out,

[57] We are glossing over a discrepancy here: according to quantum theory, such chance measures are not defined over Boolean fields—as we took them to be—but over algebraic structures of a different sort; for simplicity again, we will have to ignore this problem.

> the conditional chance of her surviving given she were to have the operation on Friday

might be different for Monday, Tuesday, and Wednesday; in that case, according to a semantics such as ours,

> If she were to have the operation on Friday she would survive

would have different truth values with reference to different times. As she says (Edgington, 2008, p. 13), "This is perhaps not a knock-down objection, but we usually think of truth values as more lasting features of our claims than probabilities."

If chances of events or propositions are only determined relative to points of time—as we will take for granted now—then a chance-based semantics for subjunctive conditionals indeed predicts that, accordingly, subjunctive conditionals have truth values relative to points of time only. But contrary to Edgington, we believe that this prediction actually gets confirmed: in the example above, the patient's condition *on Monday* might be such that if she were to have the operation on Friday she would survive, while *on Tuesday* her condition might have deteriorated to such a degree that it is no longer that case that if she were to have the operation on Friday she would survive. Instead we agree with Bovens (1998, p. 79) that "It has received little attention that the truth values of sequential counterfactuals (i.e., counterfactuals in which the antecedent event precedes the consequent event) can shift over time." Here is one of Bovens' (1998, p. 80f) examples:[58]

> Both Johann and Ludwig are inviting friends to a joint party. Johann is trying to decide whether to invite Bettina (at time t) or not. He knows that Ludwig will flip a coin tomorrow and invite Bettina if and only if heads comes up. Bettina will come to the party (at a later time t') if and only if she gets an invitation from either Johann and Ludwig. Now, Johann does invite Bettina. Suppose that he says at the time of inviting her:
>
> (SFi) Even if I were not to invite Bettina, she would still come to the party.
>
> Or suppose that after inviting Bettina, but before Ludwig's coin toss, he says:
>
> (SFii) Even if I had not invited Bettina, she would still come to the party.
>
> Ludwig's coin toss comes up heads and he invites Bettina. Suppose that, after Ludwig's coin toss but before the party, Johann says:
>
> (SFiii) Even if I had not invited Bettina, she would still come to the party.
>
> Bettina comes to the party. Suppose that, after the party, Johann says:
>
> (SFiv) Even if I had not invited Bettina, she would still have come to the party.
>
> Clearly, (SFiii) and (SFiv) are true. Even if Johann had not invited Bettina, she would still come or have to come to the party, since Ludwig

58 We state the complete example since it drives home the point very nicely. 'SF' below is Boven's abbreviation of 'semifactual', which he uses in view of the fact that the consequents in question are taken to be true. Afterwards, he constructs a similar example which does not involve semifactuals in this sense, and later on in the paper he also develops a probabilistic version of Goodman's cotenability account of subjunctive conditionals by which he suggests to analyze examples as those formally.

invited her. But what about (SFi) and (SFii)? Before Ludwig's coin toss, the coin might or might not have come up heads and so Ludwig might or might not invite Bettina to the party. Consequently, were Johann not to invite Bettina or had Johann not invited Bettina, then she might or might not come to the party. But, following Lewis, if she might not come to the party, then it is not true that she would come to the party and so, (SFI) and (SFii) are false. Hence, the truth value of the semifactual:

(SF) not-Invite(Johann, Bettina, t) $\Box\!\!\rightarrow$ Come(Bettina, t')

switches at the time of Ludwig's coin toss.

We add that these conclusions are plausible independently of Lewis' theory. Let us consider Bovens' example from the viewpoint of our theory: By t_1, t_2, t_3, t_4 we denote the points of time of utterances referred to above; hence, $t = t_1 < t_2$, t_2 is before Ludwig's coin toss, t_3 is after Ludwig's coin toss but before t', and t' is before t_4. Obviously, the conditional chance

(Chance-i) $\mathfrak{P}^{t_1}_{@}(\llbracket \text{Come(Bettina}, t')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket)$

at time t_1 (and for the actual world @) is far from 1, since it depends on the chancy event of Ludwig's tossing a hopefully fair coin. So (SFi) comes out false, as intended.

Unlike (Chance-i),

(Chance-ii) $\mathfrak{P}^{t_2}_{@}(\llbracket \text{Come(Bettina}, t')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket)$

is a chance conditional on a proposition which has absolute chance 0 at the relevant time (t_2), as the antecedent proposition describes an event at a time prior to t_2 which did not take place. For chances are understood in the way that the chance *at t* of a proposition that is solely about an event at a time *before t* is either 1 or 0, depending on whether that event took place or not. For the same reason, we cannot simply apply the Ratio formula in the case of (Chance-ii) in order to calculate the conditional probability in question. However, we may argue as follows: let t'' be the time of Ludwig's invitation, so t'' is after the coin toss but before t_3; by the Multiplication Axiom for Popper functions,

$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Come(Bettina}, t')\rrbracket \cap \llbracket \text{Invite(Ludwig, Bettina}, t'')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket) =$
$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Come(Bettina}, t')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket) \cdot$
$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Invite(Ludwig, Bettina}, t'')\rrbracket \mid \llbracket \text{Come(Bettina}, t')\rrbracket \cap \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket)$.

The last conditional probability should be 1 by the story from above, since Bettina's coming in the absence of an invitation by Johann necessitates the existence of Ludwig's invitation; moreover, the consequent of the first conditional probability expression can be simplified, for Ludwig's invitation necessitates Bettina's coming by the story again. So we end up with the equation:

$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Invite(Ludwig, Bettina}, t'')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket) =$
$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Come(Bettina}, t')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket)$.

But Ludwig's invitation is causally independent of Johann's not inviting Bettina, so it should also be probabilistically independent of it, whatever the time at which the conditional chance is taken in the actual world. Therefore,

$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Invite(Ludwig, Bettina}, t'')\rrbracket \mid \llbracket \text{not-Invite(Johann, Bettina}, t)\rrbracket) =$
$\mathfrak{P}^{t_2}_{@}(\llbracket \text{Invite(Ludwig, Bettina}, t'')\rrbracket \mid W) = \frac{1}{2}$ (roughly)

and thus

$\mathfrak{P}^{t_2}_{@}([\![\text{Come(Bettina}, t')]\!] \mid [\![\text{not-Invite(Johann, Bettina}, t)]\!]) = \frac{1}{2}$ (roughly).
It follows that (SFii) is false, too, as expected.

For t_3 and t_4 it is sufficient to observe that Ludwig's invitation has already taken place then, therefore,
$\mathfrak{P}^{t_3/t_4}_{@}([\![\text{Invite(Ludwig, Bettina}, t'')]\!] \mid W) = 1$
instead of $\frac{1}{2}$, and so by analogous reasoning as before,

$$\text{(Chance-iii/iv)}\ \mathfrak{P}^{t_3/t_4}_{@}([\![\text{Come(Bettina}, t')]\!] \mid [\![\text{not-Invite(Johann, Bettina}, t)]\!])$$

is equal to 1, which means (SFiii) and (SFiv) are true, in line with Bovens' original argument. This also shows that chances conditional on a proposition which is about an event at a time t may change from one temporal reference point to the other even when both of these reference points are *later than* t.

We find that rather than being an oddity in the present context, the time-relativity of chances naturally explains the time-relativity of truth values of subjunctive conditionals: assuming that chances determine the truth values of subjunctive conditionals, since chances need some temporal reference point t_r, also the evaluation of subjunctive conditionals needs such a reference point.[59] Bovens' argument for the time-relativity of the truth values of counterfactuals was given independently of any of the considerations that drive the probabilistic semantics in this paper, which is why it gives the semantics independent support.

This leaves us with one remaining problem: In the "official" statement of the Popper function semantics, no such reference points of evaluation were mentioned, neither in the object language nor in the metalanguage. But this should not be a major worry. Compare the situation with the one of tense modalities outside of conditional contexts: if taken at face value, 'it will be the case that A' or 'it was the case that A' are true or false only relative to a point of time. Tense logic takes care of this by supplying such reference points on the metalevel. But in a more scientific context, one would instead rather make the reference point explicit in the object language, thereby replacing the previous logically incomplete expressions by logically complete sentence of the form '$A[t]$', where t is the relevant reference point; the resulting statements would thus be true or false *simpliciter* again. Now make the same move in our conditional context: originally, in natural language, the antecedent and consequent clauses of a subjunctive conditional are tensed; in the clearest cases, in which the antecedent is about an event at one time only, the reference point of

If it were the case that A, then it would be the case that B

is usually approximately the same as the antecedent time (that is, now); the reference point of

If it had been the case that A, then it would have been the case that B

is normally after the antecedent time; and the reference point of

If it were to be the case that A, then it would be the case that B

[59] This should not be too much of a surprise, as even from a Lewisian perspective the relevant similarity ordering of possible worlds has to be taken to be codetermined by pragmatic factors such as time—usually, the time at which a counterfactual is asserted.

is typically before the antecedent time.[60] By using tense expressions, this gets signalled to the communication partner, with the reference point in question normally being determined pragmatically by the time of the assertion of any of these conditionals.[61] Now let the reference point be made explicit in the object language instead: if $A[t]$ and $B[t']$ are themselves complete sentence in the logical sense about events taking place at times t and t', respectively, then tense becomes obsolete by assuming the ultimate logical form of a subjunctive conditional is

$$A[t] \mathrel{\Box\!\!\rightarrow_{t_r}} B[t']$$

where t_r is the reference point. So $A[t]$ and $B[t']$ describe events at t and t', respectively, whereas the whole conditional $A[t] \mathrel{\Box\!\!\rightarrow_{t_r}} B[t']$ is evaluated as of the temporal point of view of t_r. Once this step of logical analysis has been taken, the resulting subjunctive conditionals are complete again and hence true or false simpliciter (i.e., relative to worlds only). On the metalinguistic side, we have to assume that each possible world w comes with an infinite sequence of conditional chance measures—one for each reference point $t_r \in \mathbb{R}$—and our truth condition for subjunctive conditionals really is:

$$w \in [\![A[t] \mathrel{\Box\!\!\rightarrow_{t_r}} B[t']]\!] \text{ iff } \mathfrak{P}_w^{t_r}([\![B[t']]\!] \mid [\![A[t]]\!]) = 1.$$

Hence, Popper functions are relativized both to possible worlds and to points of time, or rather each possible world is equipped with a time-indexed family of Popper functions, and we understand the conditional probability measures that are the components of our intended Popper function models to be such time-dependent conditional chance functions. Although the truth value of $A[t] \mathrel{\Box\!\!\rightarrow} B[t']$ is dependent on the reference point t_r being supplied, certain patterns of invariance still follow from general properties of chance: In particular, if $A[t]$ is true at w, then the truth value of $A[t] \mathrel{\Box\!\!\rightarrow} B[t']$ relative to t_r is simply the truth value of $B[t']$ as long as $t_r > t, t'$ (as we will explain in detail at the beginning of section 1 of part B), for then the chances in question will be crisp. Furthermore, independent of whether $A[t]$ is true or false at w, the truth value of $A[t] \mathrel{\Box\!\!\rightarrow} B[t']$ at w should be be one and the same for all reference points t_r that are later than t, t' as long as there is an event which actually took place at w, such that whatever gets described by $A[t]$ is causally independent of that event, and where that event together with what is described by $A[t]$ would cause the event described by $B[t']$ to happen with very high probability (we have exploited this feature in our analysis of Bovens' example above, and we will do so again in Example 3 of Section §4.7).

Furthermore, in the typical case of a counterfactual in which the antecedent describes a state or an event at a particular time t, the corresponding conditional chance of the consequent being the case given the antecedent might automatically involve a conditionalization on the complete history close to time t that still allows for the antecedent to be true, by the very nature of such time-dependent conditional chance functions. So for an $A[t]$ which

---

60 There are actually more options here if also the consequent time is taken into account: for instance, in

If it had been the case that A, then it would have to be the case that B

the reference point is after the antecedent time but typically also before (or maybe at) the consequent time, in contrast with the 'it would have been the case'-type consequent which normally is about some time before the reference point.

61 But see MacFarlane (2008) for the distinction between contexts of use and contexts of assessment of tensed claims.

describes a state or an event at t and no events at other times, the time-indexed conditional chance

$$\mathfrak{P}_w^{t_r}([\![B[t']]\!] \mid [\![A[t]]\!])$$

might, by the nature of chance—always or at least in the typical case—coincide with

$$\mathfrak{P}_w([\![B[t']]\!] \mid [\![A[t]]\!] \cap H_{<t}^w)$$

where $H_{<t}^w$ is the proposition which expresses the complete history of w close to t, though not exacty until t, such that $H_{<t}^w$ allows for a smooth transition to $A[t]$ being true without expressing anything beyond. If so, this would have to be justified by further insights into conditional chance; for the purposes of our paper, we want to leave this open, but it would of course be a natural thesis to consider. In the deviant case in which the antecedent of a counterfactual does not specify a state or an event at any one particular time, it would not be so clear anymore which antecedent-determined history proposition one would have to assume as the additional "given" argument of the time-dependent conditional chance function; but then again we often do not have a clear understanding of such deviant counterfactuals either. Maybe in such a deviant case, nothing—or rather only a tautological proposition—would get added to the antecedent proposition itself when the conditional is evaluated. In particular, if the antecedent of a subjunctive conditional is logically true, the conditional chance of its consequent given such an antecedent simply amounts to the *absolute* chance at the reference point of the consequent being true. But we will be lucky enough if we manage to make progress in the understanding of prototypical cases of counterfactuals, so questions like these shall not concern us very much in the following.[62]

We have seen what details concerning our truth condition have to be supplied, and how the logical form of subjunctive conditionals has to be clarified, in order for our probabilistic

[62] One might think initially that reference to an infinite sequence of time-indexed conditional chance measures could be avoided in the following way: let w be the possible world in which we are interested, let $H_{<t_r}^w$ be the proposition expressing the complete history of w up to (but not including) t_r; then our truth condition for subjunctive conditionals might be construed as, or at least that would be the thought,

$$w \in [\![A[t] \mathrel{\Box\!\!\rightarrow}_{t_r} B[t']]\!] \text{ iff } \mathfrak{P}_w([\![B[t']]\!] \mid [\![A[t]]\!] \cap H_{<t_r}^w) = 1.$$

Hence, Popper functions would remain to be relativized to possible worlds but not to points of time. However, it is easy to see that this is *not* an option: in the case where $t < t_r$, that is, if the counterfactual is about the past as seen from the reference point, and when also $A[t]$ is false in w, then $[\![A[t]]\!] \cap H_{<t_r}^w$ is simply identical to the contradictory proposition $\varnothing$, since every world satisfying $H_{<t_r}^w$ must satisfy $\neg A[t]$; hence $A[t] \mathrel{\Box\!\!\rightarrow}_{t_r} B[t']$ would be trivially 1 even for $\mathfrak{P}_w$ being a Popper function. Moral: The relativization of chances to a reference point is not to be understood as conditionalization on complete histories up to the reference point. This said, as Gerhard Schurz suggested to us in personal communication, one might try to save the account by including a revision operator $*$ in the antecedent proposition, as in

$$w \in [\![A[t] \mathrel{\Box\!\!\rightarrow}_{t_r} B[t']]\!] \text{ iff } \mathfrak{P}_w([\![B[t']]\!] \mid [\![A[t]]\!] * H_{<t_r}^w) = 1.$$

If $*$ satisfies the postulates of belief revision theory (see Alchourrón et al., 1985), then $*$ will still take care of the consistent case conjunctively but at the same time it would turn an inconsistent proposition into a consistent one again. Of course one would have to say more about which revision operator ought to be used here and if it still ought to be given the epistemic interpretation that it has in belief revision theory. (And maybe one should rather use a belief update operator here in the sense of Grahne, 1991; Katsuno & Mendelzon, 1992, which, however, still has an epistemic interpretation.)

semantics to embrace the time-relativity of chance and counterfactuals. For most of this paper, we will nevertheless ignore this kind of time-relativity, speaking of counterfactuals as being of the simple form $A \Box\!\!\rightarrow B$, and leaving reference points aside. But this is solely because we hope the resulting division of labor will facilitate this project just as it did in the case of Lewis' semantics: we want to focus mainly on the time-invariant features of our semantics, and we will have to get back to the temporal details—including a sound and complete system of logic for counterfactuals of the logically complete form $A \Box\!\!\rightarrow_{t_r} B$, a discussion of temporal comparability between different possible worlds, and maybe an extension of the semantics in terms of a branching time structure (cf. Thomason & Gupta, 1980)—on a different occasion. To be sure, for the remainder of this article, let it be understood that where an incomplete counterfactual in the natural language sense is asserted, the Popper function that is meant to figure on the right-hand side of its truth condition is normally not just a function of the asserted sentence but rather of the assertion itself; in other words: normally Popper functions will be constrained pragmatically by the time of assertion as being the temporal reference point. They may also be constrained by other pragmatic parameters, although one may expect chances to be quite "resilient" in that respect (using Skyrms', 1980b term again) due to their ontic basis. But similar remarks could be made for Lewis' sphere systems, and indeed apart from what we said in Section §2 we do not have to say much about the pragmatics of counterfactuals in this paper beyond what Lewis had to say. At a couple of places later on, however, we *will* return to the time relativization of chance, as it will prove necessary then.

4.4. Indeterminism and a closely related approach. We will not enter the debate at all on whether the existence of a nontrivial chance function in a world w entails that w is nondeterministic in the sense that the fundamental physical state transition laws in w are probabilistic in nature—as is the default assumption—or whether nontrivial chances at w are actually consistent with physical determinism in w; while we will continue to speak of the existence of events that have chances strictly between 0 and 1 as entailing indeterminism, this will only be so for the sake of simplicity. Indeed, as mentioned before, Loewer (2007) argues for a probabilistic understanding of subjunctive conditionals based on what he calls statistical mechanical probabilities which are given an objective nonepistemic interpretation but which he considers as *not* necessarily grounded by quantum mechanics (though this is mentioned as an option). He exemplifies this approach by means of "decision conditionals" of the form

If at t I were to decide to A then the probability of B would be x

where decisions are viewed as indeterministic relative to both the macrostate of the brain and the environment prior to, and at the moment of, the decision making, even when determinism holds. Loewer suggests that a subjunctive conditional of the form above is true if and only if[63]

$$\mathfrak{P}_@(\llbracket B \rrbracket \mid \llbracket D_A(t) \rrbracket \cap M(t)) = x$$

with $\mathfrak{P}_@$ being a uniform and time-invariant statistical probability measure, $D_A(t)$ being a sentence that describes the decision to A at t (and nothing else), and $M(t)$ being the overall macrostate at time t. Clearly, this is very close to conditional chance statements such as

$$\mathfrak{P}_@(\llbracket B \rrbracket \mid \llbracket A_t \rrbracket \cap H^@_{<t}) = x$$

---

63 We rephrase the original formulation on Loewer (2007, p. 317) very slightly.

which figured in the last subsection. Loewer then extends his account beyond decision conditionals to counterfactuals of the form 'If A_t had been true, then the chance of B would have been x' and uses his truth condition in order to avoid some challenges and counterexamples that statistical mechanics poses to Lewis' theory as found by Elga (2000). Since Loewer's truth condition for probabilistic counterfactuals is so close to ours for counterfactuals of the simple form $A \mathrel{\Box\!\!\rightarrow} B$, it may be hoped that these advantages of Loewer's theory translate to our theory, but arguing for this would require us to go much deeper into physics, and into the interpretation of probability in certain parts of physics, than we can. As Loewer says, rather than giving a metaphysical or a linguistic account, the project in his paper is to develop a scientific account of temporal asymmetries, including the temporal asymmetries that hold between the antecedent and the consequent of a typical counterfactual. We share his scientific approach to the semantics for counterfactuals.

4.5. Triviality waiting in the wings? Now that we have untangled the actual purpose of the semantics and what kind of probability is meant to be involved in its intended Popper function models, we should address some more specific worries about this probabilistic type of semantics for counterfactuals.

First, in this section, the formal concern: Since Adams' thesis (or rather Stalnaker's thesis) famously fell prey to Lewis' triviality result (see Lewis, 1976; Hájek & Hall, 1994; and the array of follow-up papers sparked by Lewis' result), it is well-known that any probabilistic semantics for conditionals which demands or presupposes for every absolute probability measure on the language of conditional logic the following requirement is bound to go straight to semantic hell: the absolute probability of a conditional is to be equal to the corresponding conditional probability of its consequent given its antecedent, where every conditional is regarded to express a uniquely determined proposition that belongs to the one Boolean field on which all of these absolute probability measures are defined, and where these absolute probability measures obey the standard axioms of probability.

In our context, demanding or presupposing in a Popper function model that

$$\mathfrak{P}_w([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid W) = \mathfrak{P}_w([\![B]\!] \mid [\![A]\!])$$

for all absolute probability measures $\mathfrak{P}_w(.|W)$ as well as

$$\mathfrak{P}^C_w([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid W) = \mathfrak{P}^C_w([\![B]\!] \mid [\![A]\!])$$

for all absolute probability measures $\mathfrak{P}^C_w(.|W)$ which arise from them by conditionalization on any C in $\mathcal{L}$—that is, postulating for all C in $\mathcal{L}$ that

$$\mathfrak{P}_w([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid [\![C]\!]) = \mathfrak{P}_w([\![B]\!] \mid [\![A]\!] \cap [\![C]\!])$$

—would simply be fatal.

Does this concern our semantics? Fortunately, it does not: the semantics only requires that

$$w \in [\![A \mathrel{\Box\!\!\rightarrow} B]\!] \text{ iff } \mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = 1$$

and hence that the *truth* of $A \mathrel{\Box\!\!\rightarrow} B$—rather than its *absolute probability*—relates to the corresponding conditional probability in the stated manner. It can be proven by straightforward construction of countless nontrivial Popper function models that this requirement is innocent.[64] Thus, the semantics does not run into Lewis' triviality result.

64 One such construction method for models with finitely many worlds is contained implicitly in the proof of Theorem 3.3 in the appendix.

However, there is still a lesson to be learnt from the triviality debate which is valuable even in our context. Assume that we wanted to amend our semantics by additional constraints on our Popper functions, such that every instance of the Import–Export scheme

$$C \mathrel{\Box\!\!\rightarrow} (A \mathrel{\Box\!\!\rightarrow} B) \leftrightarrow (C \wedge A) \mathrel{\Box\!\!\rightarrow} B$$

would turn out logically true. By the satisfaction clause for counterfactuals, this would amount to demanding

$$\mathfrak{P}_w([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid [\![C]\!]) = 1 \text{ iff } \mathfrak{P}_w([\![B]\!] \mid [\![A]\!] \cap [\![C]\!]) = 1$$

for every world $w \in W$ in every model. In other words, we would generally postulate that

$$\mathfrak{P}_w^C([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid W) = 1 \text{ iff } \mathfrak{P}_w^C([\![B]\!] \mid [\![A]\!]) = 1$$

for all $C \in \mathcal{L}$, with $\mathfrak{P}_w^C$ arising through conditionalization again by means of $\mathfrak{P}_w^C(Y|X) := \mathfrak{P}_w(Y \mid X \cap [\![C]\!])$. But this would correspond to Harper's thesis (cf. Harper, 1975) which is known to be almost equally disastrous as Stalnaker's thesis as it is subject to a form of trivialization again (see Stalnaker, 1976) whenever a conditional is presupposed to express one and the same proposition whatever the probabilistic context in which it gets evaluated.[65] So what we do find is that our semantics cannot be joined with the thesis that Import–Export is logically valid. This is of course in contrast with McGee's (1989) semantics for indicative conditionals in which a form of Import–Export is taken for granted consistently and nontrivially.

Is this a problem? We do not think so: while

$$\mathfrak{Cr}([\![A \rightarrow B]\!] \mid [\![C]\!]) = 1 \text{ iff } \mathfrak{Cr}([\![B]\!] \mid [\![A]\!] \cap [\![C]\!]) = 1$$

does seem plausible for rational conditional credence functions $\mathfrak{Cr}$ and the indicative $\rightarrow$ (see McGee, 1994; Arló-Costa, 2001 for arguments and corresponding epistemic-probabilistic models of versions of Import–Export), it is much less clear why

$$\mathfrak{P}_w([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid [\![C]\!]) = 1 \text{ iff } \mathfrak{P}_w([\![B]\!] \mid [\![A]\!] \cap [\![C]\!]) = 1$$

or

$$C \mathrel{\Box\!\!\rightarrow} (A \mathrel{\Box\!\!\rightarrow} B) \leftrightarrow (C \wedge A) \mathrel{\Box\!\!\rightarrow} B$$

should hold for conditional chance measures $\mathfrak{P}_w$ and the subjunctive $\mathrel{\Box\!\!\rightarrow}$. Indeed, Import–Export is not valid according to Lewis' semantics (and see Etlin, unpublished, for some counterexamples to subjunctive Import–Export). Moreover, as Gibbard (1981, p. 234f) has pointed out,

$$(A \supset B) \mathrel{\Box\!\!\rightarrow} (A \mathrel{\Box\!\!\rightarrow} B)$$

would be logically equivalent to

$$((A \supset B) \wedge A) \mathrel{\Box\!\!\rightarrow} B$$

[65] Here is an alternative way of seeing the problem: by defining a qualitative revision operator $*$ by means of $C * A = \{B : \mathfrak{P}_w^C([\![B]\!] \mid [\![A]\!]) = 1\}$, one could derive a trivialization result along the lines of Gärdenfors' triviality result for the belief revision version of the Ramsey Test for conditionals (cf. Gärdenfors, 1986; Gärdenfors, 1988, chap. 7). In order for this to go through, one only needs Import–Export and the additional assumption that for all C, A in $\mathcal{L}$, if C and A are consistent separately, then also $C * A$ is consistent.

by Import–Export. But the latter ought to count as logical truth on any understanding of $\Box\!\!\rightarrow$, so the former would have to be logically true as well, which sounds slightly odd. If given the additional assumption that $C \Box\!\!\rightarrow D$ logically entails $C \supset D$, even $(A \supset B) \supset (A \Box\!\!\rightarrow B)$ would then have to be logically true, and thus, under that assumption, $A \Box\!\!\rightarrow B$ would end up being logically equivalent to $A \supset B$, which would clearly be unacceptable. According to our probabilistic semantics, $A \Box\!\!\rightarrow B$ does in fact *not* logically entail $A \supset B$ (which will be discussed and defended in section 1 of part B), and thus, from this point of view, this way of reasoning does not rule out Import–Export completely but it still puts pressure on it.

We have seen that our theory does have to acknowledge that

$$\mathfrak{P}^C_@([\![B]\!] \mid [\![A]\!]) = 1 \text{ iff } \mathfrak{P}^C_@([\![A \Box\!\!\rightarrow B]\!] \mid W) = 1$$

cannot generally be true for all C, for triviality reasons (where '@' denotes the actual world). But even independent of the triviality results, there seem to be good reasons to believe that this equivalence ought *not* to hold *if seen from the viewpoint of the semantics again*. Let C be tautological: then, to be sure, 'the conditional chance of B given A is 1' sounds a lot like 'the absolute chance of $A \Box\!\!\rightarrow B$ is 1'. On the other hand, maybe 'it is maximally likely that if the match were struck it would light' *is* in fact ambiguous: the Match striking could be meant to relativize the maximal probability of the consequent which would amount to a conditional probability statement again—analogous to, for example, Kratzer's (2009) understanding of probabilistic (indicative) conditionals[66]—or it could be meant as saying that the probability of the match being flammable is maximal, which would correspond to the absolute probability statement from before. Accordingly, a (hopefully good) philosophical theory of chance and counterfactuals might convince us ultimately to drive a wedge between the two understandings. According to the probabilistic semantics defended in this paper, with C tautological, the left-hand side of the equivalence statement above expresses the *truth* of $A \Box\!\!\rightarrow B$. In contrast, its right-hand side states that $A \Box\!\!\rightarrow B$ is true in every possible world with the possible exception of a set of worlds with absolute chance 0; that is, $A \Box\!\!\rightarrow B$ is *almost necessary*, from the viewpoint of the chance measure of the actual world, since it is true in all worlds except maybe for a set of worlds that has probability 0. So at least if looked at from *within* our theory, there are equally strong reasons for keeping these two statements apart as there are for keeping apart truth and necessity for other types of sentences. Or, in other terms: if A and B do not contain $\Box\!\!\rightarrow$ themselves, then

$$\mathfrak{P}_@([\![B]\!] \mid [\![A]\!]) = \mathfrak{P}_@(\{w : B \text{ is true in } w\} \mid \{w : A \text{ is true in } w\})$$

only depends on the Popper function of the actual world, however,

$$\mathfrak{P}_@([\![A \Box\!\!\rightarrow B]\!] \mid W) = \mathfrak{P}_@(\{w : A \Box\!\!\rightarrow B \text{ is true in } w\} \mid W) = \\ \mathfrak{P}_@(\{w : \mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = 1\} \mid W)$$

also depends on the Popper functions of all other possible worlds, which is why the two should not be expected generally to be equal in value.

In any case, we find no reason for concern about the Popper function semantics in view of any of the standard triviality results on conditionals and probability.

66 We thank Paul Egré for bringing this literature to our attention.

4.6. Further worries. But there are also independent arguments to be found in the relevant literature that are directed against the probabilistic semantics for counterfactuals of Section §3 (cf. Edgington, 1995, 2004, 2008; Bennett, 2003, chap. 16).[67] In particular, Edgington (1995, p. 292f) deals with precisely our truth condition for counterfactuals when she argues (we state this in full since it concerns a crucial point):

> Prima facie, there is room for an account of objectively correct conditional thoughts. It doesn't follow that they have truth conditions. The following has been suggested[68]
>
> "If A, B" is true iff the objective probability of B given A is sufficiently high.
>
> This is not compatible with the Thesis, and is independently objectionable. (I do not object to the fact that the truth condition is vague.) Presumably, in a context, either there is some number less than 1 which is sufficiently high; or there is some number greater than 0 which is not sufficiently high; or (most likely) both. Take an example where objective probabilities are relatively easy to estimate—balls in bags, say. Call the proposed truth condition S. First suppose 0.9, say, is sufficiently high, and I am certain that the objective probability of B given A is 0.9. My degree of belief in B given A is 0.9. According to the Thesis, I am 90% confident that if A, B. But I am certain that S, hence, certain that the conditional is true. By the truth condition, I am 100% confident that if A, B. (The truth condition has the additional embarrassing consequence that the truth of "If A, B" is compatible with the truth of $A\&\neg B$.) Second, suppose some number greater than 0 is not sufficiently high—0.5 say. Suppose that I am certain that the objective probability of B given A is 0.5, and so have degree of belief 0.5 that B given A. By the Thesis, I am 50% confident that if A, B. Now I am certain that S is false, hence certain that the conditional is false. By the truth condition, I am 0% confident that if A, B. (The truth condition also has the consequence that the truth of $A\&B$ is compatible with the certain falsity of "If A, B". Not everyone minds that.)

We will split up Edgington's argumentation into parts. We will deal with her parenthetical remarks on the more or less "embarrassing" consequences of the truth condition later. Let us first turn to her main objection: we are going to argue that this objection vanishes as long as one is willing to distinguish between

[67] We will only deal with Edgington (1995) in detail since it makes the strongest case against a semantics such as ours, and the arguments in the other papers are reasonably similar. Given the aims of our article, however, it is interesting to note that Edgington (2004, 2008) mentions how easy it is for counterfactuals to be false on the standard truth conditions as an argument against the Lewisian semantics: "The truth condition of the Lewis-Goodman type are, in my view, too strong" (Edgington, 2004, p. 14). Accordingly, Bennett (2003) considers it to be one of the advantages of a "near-miss" truth condition such as ours to make many counterfactuals true which we ordinarily want to call true but which would not be true according to the standard semantics. Note that we have already dealt with one of Edgington's (2008) worries—the time-relativity worry—in our Section 4.3.

[68] Edgington cites Blackburn (1986, pp. 213–215) and an unpublished manuscript on "Conditionals" by Michael Woods.

$$(1a)\ \ \mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = \mathfrak{Cr}([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid W)$$

$$(1b)\ \ \mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = \mathfrak{Acc}_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$$

and

$$(2a)\ \ \mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = \mathfrak{Cr}([\![A \rightarrow B]\!] \mid W)$$

$$(2b)\ \ \mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = \mathfrak{Acc}_{\mathfrak{Cr}}(A \rightarrow B)$$

where in (1b) and (2b) '$\mathfrak{Acc}_{\mathfrak{Cr}}$' denotes an epistemic acceptability function again just as it did in Section §2, and where in (2a) and (2b) the symbol '$\rightarrow$' formalizes the indicative 'if . . . then'.

As we have seen in Section §2, (2b) is fairly generally accepted (though not undisputed, of course): By the original, non-Stalnakerian version of Adams' (1975) thesis for indicative conditionals, which may be regarded as the probabilistic version of the Ramsey Test (see Section §2)—or the "Thesis," in Edgington's terminology—the conditional epistemic probability of B given A of a rational agent equals the degree of acceptability that this agent attaches to the indicative conditional $A \rightarrow B$. By Lewis' triviality result, the latter cannot be equated generally with the absolute epistemic probability of $A \rightarrow B$ taken as a proposition, which is perhaps the strongest reason for following the Suppositional Theory of indicative conditionals in believing that indicative conditionals do not express propositions and hence are not true or false (or only true or false in restricted circumstances, such as when A is true). So $\mathfrak{Acc}_{\mathfrak{Cr}}$ in (2b) is *not* simply the absolute probability of an indicative conditional as determined by $\mathfrak{Cr}$—it is of quite a different nature.[69] Accordingly, while (2a) is perhaps *prima facie* plausible, it ultimately turns out to be highly suspect. But if indicative conditionals do not express propositions, there is no reason to believe that (2b) entails (2a), so this does not affect the plausibility of (2b) in any way (in fact there is then no reason to believe that 2a is even well-formed). We can grant all of this to Edgington for the sake of the argument.

How about (1b), which is simply the subjunctive version of (2b)? (Note that in Edgington's example, the conditional epistemic probability should simply be taken to coincide with the expected conditional chance for the counterfactual in question which is the value that we should actually consider.) *Prima facie*, any truth-conditional semantics of subjunctive conditionals would make it quite plausible to believe that (1a) should follow from (1b), in view of the then plausible thesis that the degree of acceptability assigned to $A \mathrel{\Box\!\!\rightarrow} B$ ought to equal the degree of belief that gets assigned to the proposition expressed by $A \mathrel{\Box\!\!\rightarrow} B$. So if (1b) were plausible in itself, (1a) would have to follow, triviality would thus be around the corner again, and our semantics would suffer from establishing a link between a plausible thesis and an inconsistent statement (given nontriviality assumptions). Presumably, this is what Edgington hints at, when she says: "This is not compatible with the Thesis." However, as we argued in detail in Section §2, $\mathfrak{Acc}_{\mathfrak{Cr}}(A \mathrel{\Box\!\!\rightarrow} B)$ should in fact *not* be assumed to coincide with $\mathfrak{Cr}([\![A \mathrel{\Box\!\!\rightarrow} B]\!]) = \mathfrak{Cr}([\![A \mathrel{\Box\!\!\rightarrow} B]\!] \mid W)$, that is, the degree of acceptability of a counterfactual is to be distinguished from its degree of belief-to-be-true: counterfactuals do have both of them as "epistemic values," but they have a different pragmatic function without either of them being redundant or implausible. It is perfectly possible that we assess subjunctive conditionals *not* in terms of epistemic probabilistic

[69] We omitted '$[\![$' and '$]\!]$' in (2b) in view of $A \rightarrow B$ not expressing a proposition according to this view.

conditionalization but by a different method, namely simple and straightforward *believing the subjunctive conditional to be true*. But that means that Edgington's objection above is questionable: In her first example, $\mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = 0.9$ while $\mathfrak{Cr}([\![A \,\Box\!\!\to B]\!] \mid W) = 1$; in her second example, $\mathfrak{Cr}([\![B]\!] \mid [\![A]\!]) = 0.5$ while $\mathfrak{Cr}([\![A \,\Box\!\!\to B]\!] \mid W) = 0$; so in each case, the first value differs from the second one, but since it is dubious if the right-hand sides of the first of two equations ought to be equal at all (with the second equation (1b) being fine in the context of Edgington's example), it is also dubious whether this is problematic in any sense. As Adams (1976, p. 17) pointed out himself, "It is not inappropriate to regard the probability 0.999 of drawing a black ball from an urn containing 99.9% black balls as something like the probabilistic disposition of the urn to yield black balls." Accordingly, in Edgington's example, the degree of belief in the truth of $A \,\Box\!\!\to B$ coincides with the degree of belief that the bag has a very high propensity of delivering a B-ball given an A-ball is drawn from it, which in turn coincides with the degree of belief that the conditional chance of B given A is very high; these degrees of belief measure something that derives from our semantic understanding of the counterfactual $A \,\Box\!\!\to B$ in terms of 'if A were the case, then with necessity/very high probability B would be the case', even though what gets measured in this way differs from what is measured by the degree of acceptability that comes attached to $A \,\Box\!\!\to B$ as well but which derives from the suppositional understanding of $A \,\Box\!\!\to B$ as being given by the Ramsey test. What a semantics such as ours *does* have to admit is that $\mathfrak{Cr}([\![B]\!] \mid [\![A]\!])$ may differ from $\mathfrak{Cr}([\![A \,\Box\!\!\to B]\!] \mid W)$, but it is unclear why this is so bad: the former is the degree by which the indicative conditional $A \to B$ is acceptable; which in some cases, as in Edgington's example, is also the degree by which the subjunctive conditional $A \,\Box\!\!\to B$ is acceptable. The latter is the degree by which the subjunctive conditional $A \,\Box\!\!\to B$ is believed to be true; and as we argued in Section §2 there are good reasons for expecting the two *not* to be identical. (And as we also showed there, for plain assertability the difference between them does not matter.) So just like in the case of (2b) and (2a), the plausible (1b) does not entail the contradictory (1a), and hence the triviality results do not affect our theory, in contrast with what Edgington thinks.

Finally, about the other consequences which Edgington rightly observes a truth condition such as ours to have: The truth of $A \,\Box\!\!\to B$ is indeed consistent with the truth of $A \wedge \neg B$—however, this is exactly as intended, as we want to allow for true counterfactuals *with exceptions*; the actual world can be an exception to its own counterfactual rule, and we are not embarrassed to say so. Moreover, the truth of $A \wedge B$ is indeed consistent with $A \,\Box\!\!\to B$ being false, but as Edgington grants, this is not that unusual to claim anyway. We shall discuss both consequences in detail when we deal with the corresponding failure of the Centering axioms in section 1 of part B.

4.7. Some advantages over Lewis' semantics. So far we have defended our semantics against a number of immediate concerns it might raise. In section 3 of part B we will give a positive argument in favor of the semantics, which is based on its compatibility of the truth of $A \,\Box\!\!\to B$ with the existence of exceptional $A \wedge \neg B$-circumstances. Section 2 in part B will conclude with the thought that the semantics has the advantage of coming close to a scientifically more respectable version of Lewis' original semantics. But already in the present section we want to point to some benefit that one gets from choosing the Popper function semantics over Lewis'. At the same time, this will show more clearly what the new semantics is like if it is put to work. As we mentioned at the beginning, our main interest is in counterfactuals which are true or false in virtue of causal affairs, and unsurprisingly our chance-based semantics proves most relevant to such counterfactuals; so the semantics has much more to say about 'If the match were struck, then it would light' than about, say, 'If Verdi and Bizet were compatriots, then Bizet would be Italian' (or French or neither). But the first of the following three examples will show that sometimes the semantics may nevertheless be useful in the evaluation of counterfactuals of noncausal sorts due to its probabilistic, rather than similarity based, format (and recall Footnote 54 for ways of incorporating purely semantic, rather than causal, information in conditional chance measures).

(1) Hájek (unpublished) (see his section 4) adapts an example that once Lewis had put forward against a Stalnaker-type Limit Assumption and uses it against Lewis' own semantics:

> If I were at least 7 feet tall, I would be precisely 7 feet tall

is clearly false. But according to Lewis' semantics it should be true, at least at first sight: for one would have to search for ways of making its antecedent true which approximate the actual world as closely as possible; but since 'at least 7 feet tall' allows for a least possible height by which I can satisfy the antecedent, and since this least height is as close as possible to my actual height within the constraints imposed by the antecedent (because I am less than 7 feet tall), the consequent should turn out to be true in the closest antecedent worlds. So Lewis' semantics faces a problem. Of course, there might be ways for the Lewisian to come back: he or she might point out that the similarity ordering in question is actually more complicated than this example makes it sound, and if the "right" similarity ordering is used the counterexample would lose its bite. But by complicating the notions of similarity or closeness, this move would also lessen the initial attractiveness and explanatory power of Lewis's semantics; it might seem as if some semantic or ontic epicycles would have to be involved in order to cover up what might point towards a more substantial theoretical defect. (Hájek, unpublished also voices such worries about "fiddling" with similarity too much.)

So much for Hájek's counterexample to Lewis' semantics. Now let us consider the same example from the viewpoint of the probabilistic semantics. In order to evaluate the counterfactual from above, we have to turn to the

> conditional chance of me being precisely 7 feet tall given that I am at least 7 feet tall.

More generally, we have to study the relevant properties of the absolute probability measure

$$\mathfrak{P}_{@}(.\,|\,[\![\text{I am at least 7 feet tall}]\!]).$$

Which height would I be likely to have if I were at least 7 feet tall? It is plausible that the chance of my height (in feet) lying within the interval $[7, 8]$ is greater than its being within the interval $[8, 9]$ which in turn is greater than the chance being anywhere within $[9, 10]$ and so on, where each of these chances is strictly less than 1, by natural laws and contingent facts about organisms in our universe. But the consequent of Hájek's counterfactual does not even specify an interval of positive length, instead it considers only a single point length—so, for all we know, the conditional chance of me being precisely 7 feet tall given that I am at least 7 feet tall should actually be *zero*. In any case, the probability will not be very high and thus Hájek's counterfactual will be false according to the semantics, as expected. We hasten to add that there is a flip side to this fact, too: for the very same reason, the conditional probability of me *not* being precisely 7 feet tall given that I am at least 7 feet tall *is* high, hence

If I were at least 7 feet tall, I would not be precisely 7 feet tall

should be true. While this might not be something that one consents to easily, it is acceptable in our eyes: If I were at least 7 feet tall, then (*apart from a single exceptional case*) I would not be precisely 7 feet tall. Or, more colloquially: If I were at least 7 feet tall, then *surely* I would not be precisely 7 feet tall, where in our case 'surely' would not be given an epistemic connotation.

(2) In the following, in some sense, related example, time is going to take over the role of height in the previous example.[70] We take Bennett's version (cf. Bennett, 2003, p. 219f) of the story[71]:

> My coat was not stolen from the restaurant where I left it. There were two chances for theft—two times when relevant indeterminacies or small miracles could have done the trick. They would have involved different potential thieves; and the candidate for the later theft is a rogue who always sells his stuff to a pawnbroker named Fence.

If "maximal similarity" in Lewis' semantics involved anything like a "maximal delay of divergence from the actual course of events," then

If my coat had been stolen from the restaurant, it would now be in Fence's shop

should be true according to the story above. Consequently, a latest-admissible-temporal-fork constraint is not what the Lewisian should settle for, but again this puts some pressure on evaluating counterfactuals on the basis of a plausible notion of similarity of worlds.

In our semantics, the coat example is not trivial to handle either, in so far as the antecedent does not specify an event at a unique point of time, which—as we pointed out in Section §4.2—makes it harder to compute the relevant probabilities. But there is still a way of treating the last example semantically that is plausible from a probabilistic point of view. According to the story above, the world offers two salient possibilities of how the coat could have been stolen: it could have been stolen by thief 1 at time t or by thief 2 at time $t + \Delta t$; let A^1_t and $A^2_{t+\Delta t}$ express propositions which describe precisely these two possible events, respectively. So, simplifying a bit, what we are interested in is the conditional chance

$$\mathfrak{P}_{@}(\llbracket \text{My coat is in Fence's shop (at some } t' > t + \Delta t) \rrbracket \mid \llbracket A^1_t \vee A^2_{t+\Delta t} \rrbracket).$$

For all we know, this conditional chance is not even close to 1, since the story did not at all make it sound *far more likely* that, given the coat had been stolen at all, the second thief would have done it. Hence, the corresponding counterfactual is false. Again, this is just as intended.

(3) We turn to one final example. Edgington (2004, p. 12) describes the following plane crash event:

> I am driving to the airport to catch a nine o'clock flight to Paris. The car breaks down on the motorway. I sit there, gnashing my teeth, waiting for the breakdown service. Nine o'clock passes: I've missed my flight. More time passes. 'If I had caught the plane, I would have been half way to Paris by now', I say to the repairman who eventually shows up.

70 We would like to thank Thomas Müller for a very interesting discussion on this point.

71 Bennett cites Donald Nute on this, who in turn learned it from John Pollock.

> 'Which flight were you on?' he asks. I tell him. 'Well you're wrong', he says. 'I was listening to the radio. It crashed. If you had caught that plane, you would be dead by now.'

As Edgington points out, this story creates difficulties for more or less all well-known theories of counterfactuals, as do similar examples which were put forward by Sidney Morgenbesser in the 1970s; the reason is simply that what actually happened *after* the time of the envisioned fork that is needed in order to make the antecedent true counterfactually turns out to be still relevant for the consequent. Edgington (2004) and Schaffer (2004) show that it is possible to tackle these difficulties by adding clauses on causal independence to the standard theories for counterfactuals; prior to them, Kvart (1986, 1992, 2001, unpublished) had done so already in his probabilistic variant of the metalinguistic theory of counterfactuals.

But within the semantic theory of this paper, no special clauses on causal independence need to be added really, just general considerations on probabilistic versus causal dependence suffice. Using the terminology of Section 4.3, it is clear that the reference point t_r of the repairman's utterance is some time after the plane crash. So evaluating

> If you had caught that plane, you would be dead by now

as described in the story is nothing but evaluating

> You catch the plane at t $\Box\!\!\rightarrow_{t_r}$ you are dead by t'

with $t < t' < t_r$ chosen respectively. But that counterfactual is true at the actual world if and only if

$$\mathfrak{P}^{t_r}_{@}(\llbracket\text{you are dead by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) = 1.$$

The chance expression in question is conditional on a proposition which has absolute chance 0 at t_r, since you had missed the plane by t_r. Once again, therefore, one cannot use the Ratio formula in order to calculate the corresponding conditional probability. But analogously to the reasoning in Section 4.3, one may instantiate the Multiplication Axiom for Popper functions and derive:

$\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \cap \llbracket\text{you are dead by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) =$
$\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) \cdot$
$\mathfrak{P}^{t_r}_{@}(\llbracket\text{you are dead by } t'\rrbracket \mid \llbracket\text{the plane crashes by } t'\rrbracket \cap \llbracket\text{you catch the plane at } t\rrbracket)$.

The last conditional probability must be close to 1 by the story above; the consequent of the first conditional probability expression can be simplified, as you being dead by t' almost necessitates the plane crash having taken place by t' (other possible causes of your death being extremely unlikely), according to the story again. This yields (roughly):

$\mathfrak{P}^{t_r}_{@}(\llbracket\text{you are dead by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) =$
$\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket)$.

But the plane crashing by t' should be more or less causally independent of you catching the plane at t, so the former should also be probabilistically independent of the latter, at whatever time. Therefore,

$\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) =$
$\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \mid W) = 1$

and hence also

$\mathfrak{P}^{t_r}_{@}(\llbracket\text{you are dead by } t'\rrbracket \mid \llbracket\text{you catch the plane at } t\rrbracket) = 1$.

It follows that

You catch the plane at t $\Box\!\!\to_{t_r}$ you are dead by t'

is true, which is what Edgington argues for on independent grounds.[72] If one does not believe the probabilistic independence claim, since one's presence on the plane might have changed the weight distribution in the plane and so on significantly, then presumably one does not hold the original counterfactual true either.

While the argument above proceeds by invoking the notion of causal independence again, it is not the case that any previous formulation of our truth condition needs amendment of any sort or that our truth condition was designed to handle cases like these (or Morgenbesser's) from the start. The plain 'a conditional chance close to 1 amounts to truth' idea suffices once it has been made precise enough, since causal independence translates into probabilistic independence simply by the nature of chance. Finally, there is also something we can offer to those who *do* regard the counterfactual in question as false in view of some "openness of the future" considerations as seen from the viewpoint of entering the plane: for if the reference point is chosen to be the point of your (counterfactually) stepping onto the plane, then the counterfactual *does* turn out to be false, by the probabilistic semantics again, because $\mathfrak{P}^{t_r}_{@}(\llbracket\text{the plane crashes by } t'\rrbracket \mid W)$ is not equal or even close to 1 anymore for $t' > t_r$. So the debate on these examples might actually oscillate tacitly between two possible reference points (even though we take the story above to intend the reference point to be the later one).

As mentioned before in Section 4.1, in our view not too much ought to be read into examples from commonsense reasoning and conversation if taken in isolation from more general theoretical considerations. But we hope the examples of this section demonstrate at least how the probabilistic semantics is meant to operate and why it is worth pursuing such a probabilistic semantics for counterfactuals at all. Also, a spoonful of sugar may help the medicine go down that we will turn to at the beginning of part B when we will deal with the Centering axioms' failing to be logically true in our probabilistic semantics. [TO BE CONTINUED BY PART B.]

§5. Acknowledgements. For their invaluable comments we want to thank Horacio Arló-Costa, Johan van Benthem, Dorothy Edgington, David Makinson, Brian Skyrms, Tim Williamson, Steve Yablo, as well as Frank Arntzenius, Alexander Bird, Franz Dietrich, Richard Dietz, Igor Douven, Paul Egré, David Etlin, André Fuhrmann, Alan Hájek, Vincent Hendricks, Leon Horsten, James Ladyman, Barry Loewer, Thomas Müller, Daniel Nolan, Rohit Parikh, Richard Pettigrew, Richard Bradley, Jan-Willem Romeijn, Daniel Rothschild, Hans Rott, Gerhard Schurz, Wolfgang Spohn, Robbie Williams, and members of the audiences at talks in Bristol, Copenhagen, Rotterdam, Oxford, Konstanz, Irvine, Manchester, Utrecht, Frankfurt, Düsseldorf, Princeton, Regensburg, London, Sheffield. We would like to thank the Leverhulme Trust and the Alexander von Humboldt Foundation for their generous support of this work.

§6. Appendix: Proofs This is the proof of the result in Section §2:

Proof of Theorem 2.1.

- Let $0 < \alpha \leq 1$. For $\epsilon > 0$, choose $\delta = \epsilon \cdot \alpha$. Then we have:

[72] For the same reason, also the suppositional degree of acceptability for that counterfactual—as given by its associated expected conditional chance—must be 1 or very close to 1. Again the conditional chances in question have to be taken relative to the reference point t_r, which is way after the plane crash.

$1-\delta \leq \sum_w \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) = \sum_{w:\mathfrak{P}_w(B|A)\geq 1-\alpha} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) + \sum_{w:\mathfrak{P}_w(B|A)<1-\alpha} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) \leq \sum_{w:\mathfrak{P}_w(B|A)\geq 1-\alpha} \mathfrak{Cr}(\{w\}) \cdot 1 + \sum_{w:\mathfrak{P}_w(B|A)<1-\alpha} \mathfrak{Cr}(\{w\}) \cdot (1-\alpha) = \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\}) + \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) < 1-\alpha\}) \cdot (1-\alpha) = 1 - \alpha \cdot \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) < 1-\alpha\}) = 1-\alpha\cdot(1-\mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\})) = 1-\alpha+\alpha\cdot\mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\})$. Therefore, $\frac{\alpha-\delta}{\alpha} \leq \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\})$ (here we use that $0 < \alpha$), that is, $1 - \frac{\delta}{\alpha} = 1 - \epsilon \leq \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\})$, which is what we needed to prove.

- Let $0 \leq \alpha \leq 1$. For $\epsilon > 0$, choose $\delta = \epsilon$. It follows:
  Since by assumption $1-\delta \leq \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\})$, we may conclude $\sum_w \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) = \sum_{w:\mathfrak{P}_w(B|A)\geq 1-\alpha} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) + \sum_{w:\mathfrak{P}_w(B|A)<1-\alpha} \mathfrak{Cr}(\{w\}) \cdot \mathfrak{P}_w(B|A) \geq \sum_{w:\mathfrak{P}_w(B|A)\geq 1-\alpha} \mathfrak{Cr}(\{w\}) \cdot (1-\alpha) = \mathfrak{Cr}(\{w : \mathfrak{P}_w(B|A) \geq 1-\alpha\}) \cdot (1-\alpha) \geq (1-\delta)(1-\alpha) = (1-\epsilon)(1-\alpha)$, as to be shown. □

Now we turn to the proofs of the theorems in Section §3:

Proof of Theorem 3.3. Soundness may be proven directly. Completeness follows from Lewis' completeness proof for V in Lewis (1973b) together with the observation that the Lewisian semantics can be used to define a Popper function if only W is a finite set, which is not an actual restriction, since by Lewis' results every formula that is not logically true in his semantics has a finite countermodel. Here is the proof sketch: If $\langle W, (\mathfrak{S}_w)_{w\in W}, [\![.]\!]\rangle$ is a Lewisian spheres model with W being finite, and if $w \in W$, let $\mathfrak{A} = \{[\![A]\!] \mid A \in \mathcal{L}\}$, for $A, B \in \mathcal{L}$ let S be the least sphere in $\mathfrak{S}_w$ that includes A-worlds (which exists by the finiteness of W) if there is an A-permitting sphere in $\mathfrak{S}_w$ at all, and in that case set $\mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = \frac{card(S\cap[\![A\wedge B]\!])}{card(S\cap[\![A]\!])}$; otherwise, if A is impossible from the viewpoint of w, let $\mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) = 1$. It follows that any such $\mathfrak{P}_w$ is well-defined and indeed a Popper function on $\mathfrak{A}$, and the resulting $\langle W, \mathfrak{A}, (\mathfrak{P}_w)_{w\in W}, [\![.]\!]\rangle$ is a Popper function model, which can be used as a countermodel for formulas that are not theorems of V by means of the fact that if w does not satisfy $A \;\Box\!\!\rightarrow B$ in the Lewisian countermodel, then $\mathfrak{P}_w([\![B]\!] \mid [\![A]\!]) < 1$ by the construction above, and hence w does not satisfy $A \;\Box\!\!\rightarrow B$ in our probabilistic countermodel either. □

Proof of Theorem 3.5. Soundness may be proven by exactly the same methods as the ones used in Hawthorne & Makinson (2007) for analogous purposes in their context of nonmonotonic consequence relations. □

BIBLIOGRAPHY

Achinstein, P. (2001). *The Book of Evidence*. Oxford, UK: Oxford University Press.

Adams, E. W. (1965). On the logic of conditionals. *Inquiry*, **8**, 166–197.

Adams, E. W. (1975). *The Logic of Conditionals: An Application of Probability to Deductive Logic*. Synthese Library 86. Dordrecht, The Netherlands: Reidel.

Adams, E. W. (1976). Prior probabilities and counterfactual conditionals. In Harper, W. L., and Hooker, C. A., editors. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht, The Netherlands: Reidel, pp. 1–21.

Albert, D. Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, **50**, 510–530.

Arló-Costa, H. (2001). Bayesian epistemology and epistemic conditionals: On the status of the export-import laws. *Journal of Philosophy*, XCVIII, 555–598.

Arló-Costa, H., & Parikh, R. (2005). Conditional probability and defeasible inference. *Journal of Philosophical Logic*, **34**, 97–119.

Baltag, A., & Smets, S. (2008). Probabilistic dynamic belief revision. *Synthese*, **165**, 179–202.

Barker, S. (1999). Counterfactuals, probabilistic counterfactuals and causation. *Mind*, **108**, 427–469.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford, UK: Clarendon Press.

Blackburn, S. (1986). How can we tell whether a commitment has a truth condition? In Travis, C., editors. *Meaning and Interpretation*. Oxford, UK: Basil Blackwell, pp. 201–232.

Bovens, L. (1998). Sequential counterfactuals, cotenability and temporal becoming. *Philosophical Studies*, **90**, 79–101.

Bratman, M. (1992). Practical reasoning and acceptance in a context. *Mind*, **101**, 1–15.

Easwaran, K. (2005). What conditional probability must (almost) be. *Paper presented at the Formal Epistemology Workshop 05*, Austin.

Edgington, D. (1995). On conditionals. *Mind*, **104**, 235–329.

Edgington, D. (1996). Vagueness by degrees. In Keefe, R., and Smith, P., editors. *Vagueness. A Reader*. Cambridge, MA: The MIT Press, pp. 294–316.

Edgington, D. (1997). Truth, objectivity, counterfactuals and Gibbard. *Mind*, **106**, 107–116.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In Dowe, P., and Noordhof, P., editors. *Cause and Chance: Causation in an Indeterministic World*. London: Routledge, pp. 12–27.

Edgington, D. (2008). Counterfactuals. *Proceedings of the Aristotelian Society*, **108**, 1–21.

Eells, E., & Skyrms, B., editors. (1994). *Probability and Conditionals. Belief Revision and Rational Decision.* Cambridge, UK: Cambridge University Press.

Elga, A. (2000). Statistical mechanics and the asymmetry of counterfactuals. *Philosophy of Science*, **68**, 313–324.

Etlin, D. (unpublished). Modus ponens revisited. Unpublished draft.

Gärdenfors, P. (1986). "Belief Revisions and the Ramsey Test for Conditionals." *Philosophical Review*, **95**, 81–93.

Gärdenfors, P. (1988). *Knowledge in Flux.* CambridgeMA: The MIT Press.

Gibbard, A. (1981). Two recent theories of conditionals. In Harper, W., Stalnaker, R., and Pearce, G., editors. *Ifs. Conditionals, Belief, Decision, Chance, and Time*. Dordrecht, The Netherlands: Reidel, pp. 211–247.

Gillies, T. (2007). Counterfactual scorekeeping. *Linguistics and Philosophy*, **30**, 329–360.

Grahne, G. (1991). Updates and counterfactuals. In Allen, J. A., Fikes, R., and Sandewall, E., editors. *Principles of Knowledge Representation and Reasoning.* San Mateo, CA: Morgan Kaufmann, pp. 269–276.

Hájek, A. (2003). What conditional probability could not be. *Synthese*, **137**, 273–323.

Hájek, A. (unpublished). Most counterfactuals are false. Unpublished draft.

Hájek, A., & Hall, N. (1994). The hypothesis of the conditional construal of conditional probability. In Eells, E., and Skyrms, B., editors. *Probability and Conditionals. Belief Revision and Rational Decision*. Cambridge, UK: Cambridge University Press, pp. 75–111.

Harper, W. L. (1975). Rational belief change, Popper functions and counterfactuals. *Synthese*, **30**, 221–262. Reprinted in Harper & Hooker (1981, 73–112).

Harper, W. L., & Hooker, C. A. (1976). *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht, The Netherlands: Reidel.

Harper, W., Salnaker, R., & Pearce, G. (1981). *Ifs. Conditionals, Belief, Decision, Chance, and Time*. Dordrecht, The Netherlands: Reidel.

Hawthorne, J. (1996). On the logic of nonmonotonic conditionals and conditional probabilities. *Journal of Philosophical Logic*, **25**, 185–218.

Hawthorne, J. (2005). "Chance and Counterfactuals." *Philosophy and Phenomenological Research*, **70**, 396–405.

Hawthorne, J. (2007). "Nonmonotonic Conditionals that Behave Like Conditional Probabilities Above a Threshold." *Journal of Applied Logic*, **5**, 625–637.

Hawthorne, J., & Makinson, D. (2007). The quantitative/qualitative watershed for rules of uncertain inference. *Studia Logica*, **86**, 247–297.

Humphreys, P. (1985). Why propensities cannot be probabilities. *Philosophical Review*, **94**, 557–570.

Joyce, J. (1999). *The Foundations of Causal Decision Theory*. Cambridge, UK: Cambridge University Press.

Katsuno, H., & Mendelzon, A. (1992). On the difference between updating a knowledge base and revising it. In Gärdenfors, P., editor. *Belief Revision*. Cambridge, UK: Cambridge University Press, pp. 183–203.

Kaufmann, S. (2005). Conditional predictions. *Linguistics and Philosophy*, **28**, 181–231.

Kratzer, A. (1981). Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, **10**, 201–216.

Kratzer, A. (2009). Modals and conditionals again. Chapter 4 of unpublished book manuscript, to be published by Oxford University Press.

Kvart, I. (1986). *A Theory of Counterfactuals.* Indianapolis, IN: Hackett.

Kvart, I. (1992). Counterfactuals. *Erkenntnis*, **36**, 1–41.

Kvart, I. (2001). The counterfactual analysis of cause. *Synthese*, **127**, 389–427.

Kvart, I. (unpublished). The causal-process-chance-based analysis of counterfactuals." Unpublished manuscript.

Levi, I. (1977). Subjunctives, dispositions and chances. *Synthese*, **34**, 423–455.

Levi, I. (1984). *Decisions and Revisions*. Cambridge, UK: Cambridge University Press.

Levi, I. (1996). *For the Sake of the Argument*. Cambridge, UK: Cambridge University Press.

Lewis, D. K. (1973a). Counterfactuals and comparative possibility.*Journal of Philosophical Logic*, **2**, 418–446. Reprinted in Harper et al. (1981, pp. 57–85) and Lewis Lewis (1986, p. 3–31).

Lewis, D. K. (1973b). *Counterfactuals*. Oxford, UK: Blackwell.

Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, **85**, 297–315. Reprinted in Harper et al. (1981, pp. 129–147) and Lewis (1986, pp. 133–152).

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In Jeffrey, R. C., editor. *Studies in Inductive Logic and Probability*, Vol. 2. Berkeley, CA: University of California Press, pp. 263–293. Reprinted in Harper et al. (1981, pp. 267–297) and Lewis (1986, pp. 83–113).

Lewis, D. K. (1981). Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, **10**, 217–234.

Lewis, D. K. (1986). *Philosophical Papers*, Vol. 2. Oxford, UK: Oxford University Press.

Loewer, B. (2007). Counterfactuals and the second law. In Price, H., and Corry, R., editors. *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford, UK: Oxford University Press, pp. 293–326.

MacFarlane, J. (2008). Truth in the garden of forking paths. In Kölbel, M., and García-Carpintero, M., editors. *Relative Truth*. Oxford, UK: Oxford University Press, pp. 81–102.

Makinson, D. (unpublished). Conditional probability in the light of qualitative belief change. Unpublished manuscript.

McGee, V. (1989). Conditional probabilities and compounds of conditionals. *The Philosophical Review*, **98**, 485–541.

McGee, V. (1994). Learning the impossible. In Eells, E., and Skyrms, B., editors. *Probability and Conditionals. Belief Revision and Rational Decision.* Cambridge, UK: Cambridge University Press, pp. 177–199.

Paglieri, F. (2009). Acceptance as conditional disposition. In Hieke, A., and Leitgeb, H., editors. *Reduction. Between the Mind and the Brain*. Frankfurt a.m., Germany: Ontos Press, pp. 29–49.

Paris, J. B., & Simmonds, R. (2009). *Review of Symbolic Logic*, **2**, 298–309.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligence Systems*. San Mateo, CA: Morgan Kaufmann.

Pearl, J. (1994). From Adams' conditionals to default expressions, causal conditionals, and counterfactuals. In Eells, E., and Skyrms, B., editors. *Probability and Conditionals. Belief Revision and Rational Decision.* Cambridge, UK: Cambridge University Press, pp. 47–74.

Pearl, J. (1996). Causation, action and counterfactuals. In Shoham, Y., editor. *TARK'96. Proceedings of the 6th conference on Theoretical Aspects Of Rationality And Knowledge*. Los Angeles, CA: Morgan Kaufmann, pp. 51–73.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press.

Popper, K. R. (1968). *The Logic of Scientific Discovery*. Revised edition. London: Hutchinson.

Ramsey, F. P. (1929). General propositions and causality. In Braithwaite, R. B., editor. *The Foundations of Mathematics and other Logical Essays*. London: Kegan Paul, Trench & Trubner, pp. 237–255.

Resnik, D. B. (1992). Convergent realism and approximatetruth. In *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1. Contributed Papers. Chicago, IL: The University of Chicago Press, pp. 421–434.

Roeper, P., & Hugues, L. (1999). *Probability Theory and Probability Logic*. Toronto, Canada: University of Toronto Press.

Schaffer, J. (2004). Counterfactuals, causal independence and conceptual circularity. *Analysis*, **64**, 299–309.

Schurz, G. (1998). Probabilistic semantics for Delgrande's conditional logic and a counterexample to his default logic. *Artificial Intelligence*, **102**, 81–95.

Schurz, G. (2001). What is 'Normal'? An evolution-theoretic foundation of normic laws and their relation to statistical normality. *Philosophy of Science*, **68**, 476–497.

Schurz, G. (2005). Non-monotonic reasoning from an evolution-theoretic perspective: Ontic, logical and cognitive foundations. *Synthese*, **146**, 37–51.

Seidenfeld, T., Schervish, M. J., & Kadane, J. B. (2001). Improper regular conditional distributions. *The Annals of Probability*, **29**, 1612–1624.

Smith, M. (2007). Ceteris paribus conditionals and comparative normalcy. *Journal of Philosophical Logic*, **36**, 97–121.

Skyrms, B. (1980a). The prior propensity account of subjunctive conditionals. In Harper, W., Stalnaker, R., and Pearce, G., editors. *Ifs. Conditionals, Belief, Decision, Chance, and Time*. Dordrecht, The Netherlands: Reidel, pp. 259–265.

Skyrms, B. (1980b). *Causal Necessity*. New Haven, CT: Yale University Press.

Skyrms, B. (1984). *Pragmatics and Empiricism*. New Haven, CT: Yale University Press.

Skyrms, B. (1994). Adams conditionals. In Eells, E., and Skyrms, B., editors. *Probability and Conditionals. Belief Revision and Rational Decision.* Cambridge, UK: Cambridge University Press, pp. 13–26.

Skyrms, B., Bell, G. D., & Woodruff, P. (1999). Theories of counter-factual and subjunctive conditionals in contexts of strategic interaction. *Research in Economics*, **53**, 275–291.

Spohn, W. (1988). Ordinal conditional functions. A dynamic theory of epistemic states. In Harper, W. L., and Skyrms, B., editors. *Causation in Decision, Belief Change, and Statistics*, Vol. 2. Dordrecht, The Netherlands: Kluwer, pp. 105–113.

Stalnaker, R. C. (1968). A Theory of conditionals. In *Studies in Logical Theory. American Philosophical Quarterly Monograph Series 2*. Oxford, UK: Blackwell, pp. 98–112. Reprinted in Harper et al. (1981, pp. 41–55).

Stalnaker, R. C. (1970). Probability and conditionals. *Philosophy of Science*, **37**, 64–80. Reprinted in Harper et al. (1981, pp. 107–128).

Stalnaker, R. C. (1976). Letter to Harper. In Harper, W. L., and Hooker, C. A. (1976). *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht, The Netherlands: Reidel, pp. 113–116.

Stalnaker, R. C. (1984). *Inquiry*. Cambridge, MA: The MIT Press.

Stalnaker, R. C., & Jeffrey, R. C.. (1994). Conditionals as random variables. In Eells, E., and Skyrms, B., editors. *Probability and Conditionals. Belief Revision and Rational Decision.* Cambridge, UK: Cambridge University Press, pp. 31–46.

Thomason, R., & Gupta, A. (1980). A theory of conditionals in the context of branching time. *Philosophical Review*, **89**, 65–90. Reprinted in Harper et al. (1981, pp. 299–322).

Unwin, N. (2007). *Aiming at Truth*. New York, NY: Palgrave Macmillan.

van Benthem, J., Gerbrandy, J., & Kooi, B. (2009b). Dynamic update with probabilities. *Studia Logica*, **93**, 67–96.

van Benthem, J., Girard, P., & Roy, O. (2009a). Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, **38**, 83–125.

van Fraassen, B. C. (1976). Representation of conditional probabilities. *Journal of Philosophical Logic*, **5**, 417–430.

van Fraassen, B. C. (1980). *The Scientific Image*. Oxford, UK: Oxford University Press.

Veltman, F. (2005). Making counterfactual assumptions. *Journal of Semantics*, **22**, 159–180.

von Fintel, K. (2001). Counterfactuals in a dynamic context. In Kenstowicz, M., editor. *Ken Hale: A Life in Language*. Cambridge, MA: The MIT Press, pp. 123–152.

Woodruff, P. W. (1999). Partitions and conditionals. *Journal of Philosophical Logic*, **28**, 113–128.