In Defence Of A Naïve Conditional Epistemology

Draft: 20/06/2012, Updated: 13/05/2013

Abstract

Numerous triviality results have been directed at a collection of views that identify the probability of a conditional sentence with the conditional probability of the consequent on its antecedent. In this paper I argue that this identification makes little sense if conditional sentences are context sensitive. The best alternative, I argue, is a version of the thesis which states that someone whose total evidence is E should have the same conditional credence in B given A as they have credence in the proposition expressed by 'if A then B' in a context where E is salient. The biggest challenge to this thesis comes from the 'static' triviality arguments developed by Stalnaker, and Hájek and Hall. It is argued that these arguments rely on invalid principles of conditional logic and that the thesis is consistent with a reasonably strong logic that does not include the principles in question.

Imagine that a coin has been selected at random from a bag containing three 20p coins, five 5p coins, six 1p coins, and four 2p coins.[1] Given that no coin has a better chance of being selected than any other, what is the probability that

(i) The selected coin is 5p if it's silver?

(ii) The selected coin is 20p if it's silver?

(iii) The selected coin is 1p if it's copper?

(iv) The selected coin is 2p if it's copper?

Intuitively the answer to these questions are: (i) $\frac{5}{8}$ (ii) $\frac{3}{8}$ (iii) $\frac{6}{10}$ and (iv) $\frac{4}{10}$. For example, to work out (i), i.e. to calculate the probability that if the coin is silver it's 5p, we consider the proportion of silver coins which are 5p coins. Since five out of seven silver coins are 5p coins we conclude that $\frac{5}{8}$ is the answer to (i).

This simple account cannot be the whole story. For example not only does the answer to questions (i-iv) depend on what your evidence is, but sometimes *which question is being asked* depends on your evidence; indicative conditionals are notoriously context sensitive.

[1] For those not familiar with British currency the first two kinds of coins are silver coloured while the latter two are copper coloured.

To see this we might consider a variant of the 'Sly Pete' examples described by Gibbard [10]. Suppose that, unbeknownst to Alice, the coin is a 1p coin, and, furthermore, she has been told that the selected coin is not a 20p coin. Alice reasons, quite correctly, that

$$\text{If it's silver then it's a 5p coin} \tag{1}$$

On the other hand Bob has been told only that the selected coin is not a 5p coin. Bob reasons, again quite correctly, that

$$\text{If it's silver then it's a 20p coin} \tag{2}$$

Some theorists have taken examples such as these to show that conditionals do not have truth conditions at all (indeed this is Gibbards own response.) However, assuming conditionals do have truth conditions, and assuming that the propositions asserted by (1) and (2) in a single context are inconsistent with one another, it seems inevitable that conditionals must exhibit some degree of context sensitivity. If Alice's utterance of (1) and Bob's utterance of (2) both express true propositions, it follows that which proposition is expressed by an utterance of (1) or (2) is sensitive to the context of utterance. In particular, since the only relevant difference between the context of Alice's utterance and of Bob's is the total evidence they have, it follows that the proposition expressed by an utterance of (1) or (2) may depend on a contextually salient piece of evidence.[2]

Bearing in mind that what question you are asking in (i-iv) may depend on your current evidence, it is natural to ask: what general principle guides our judgments in (i-iv)? If the reported judgments in (i-iv) are in fact correct then, since they were calculated in a uniform manner, one would expect there to be some general explanation of them.

Probability judgments involving conditionals, such as those reported in (i-iv), are determined by two pieces of evidence: one piece of evidence determines which proposition is asserted by an utterance of the conditional, the other determines the probability function with which we make the actual judgments of probability; the former determines the proposition to be evaluated, and the latter determines how probable that proposition is. My proposal, then, is that when these two peices of evidence are identical then the probability of the conditional and the conditional probability coincide. In other words, if one's total evidence is Γ then the evidential probability of a conditional that's evaluated when the evidence Γ is salient is identical to the conditional evidential probability of the consequent on the antecedent. The conditional probability of B given A, written $Pr(B \mid A)$, is determined by the ratio $\frac{Pr(A \wedge B)}{Pr(A)}$.[3]

[2] I do not want to limit myself to the claim that an utterance of a conditional always depends on the utterer's evidence – it may be some amalgamation of the evidence of the participants of the conversation or some other contextually salient piece of evidence.

[3] We might compare this to van Fraassen's denial of 'metaphysical realism' in [37] in which the interpretation of the conditional itself depends on the probability function you are evaluating it with respect to. However, unlike van Fraassen's metaphysical claim, the *contextual* dependence I am positing is quite respectable.

Although a simple and appealing theory of the epistemology of conditionals, we shall see that many philosophers argue, citing triviality results, that this style of explanation must be incorrect – at least, if conditionals of this sort are to have truth conditions at all (see [12] for a representative example.) This position, however, raises a whole host of difficult questions. For example: what should we make of our intuitive probability judgments about conditionals? One reaction would be to take these highly theoretical arguments as reasons to reject these judgments. Needless to say, I think this would be unwise – triviality results do not cast doubt on particular probability judgments such as those reported in (i-iv). They merely refute a general theory that predicts those judgments – the judgments themselves do not imply the refuted theory. Furthermore, if the answers I listed to (i-iv) are not correct then this theorist owes us an answer to the question: what *are* the correct answers to these particular questions? Another reaction, my own, would be to accept judgments such as (i-iv) and propose an alternative theory that predicts them. In my opinion no-one has yet given a satisfactory answer to these questions which does not invoke the relation between conditionals and conditional probabilities.

Let us write $A \to_\Gamma B$ for the proposition expressed by an utterance of a conditional relative to evidence Γ, whose antecedent expresses A and consequent expresses B. The main technical result of this paper is that the below principle, CP – a formal rendering of the principle outlined above – is consistent, and holds for a large range of non-trivial probability functions.

CP $Pr(\psi \mid \Gamma \cup \{\phi\}) = Pr(\phi \to_\Gamma \psi \mid \Gamma)$.

Here Γ ranges over sets of propositions which could, in some possible world, be some agent's total evidence. For now it will not matter if we allow Γ to be any consistent set of propositions whatsoever. Pr ranges over rational initial probability functions, sometimes called 'priors' or 'ur-priors' – the rational credences of a completely uninformed agent. There is nothing particularly mysterious about ur-priors – in order to have informed rational credences at all there has to be credences it would be permissible to have if you had no evidence whatsoever. I shall make the standard assumption that ur-priors are regular in the sense that no possibility is ruled out by the function *a priori* – formally speaking, that the function takes value 0 only on the inconsistent proposition. The evidential probability of an agent at a given time can therefore be identified with the result of conditioning her ur-prior on her total evidence at that time, $Pr(\cdot \mid \Gamma)$. This would not make sense if the agent is able to initially rule out some of her later evidence *a priori*, which is the reason regularity must be assumed. CP is compatible with the thesis that there is exactly one prior, E, representing the evidential probability prior to investigation (see Williamson [38].) Nothing I say in this paper relies on there being only one rational prior, so my defence of the thesis is left at a general level. It is important to note that sometimes 'prior probability', or 'initial probability', is used to mean the credence of an informed rational agent before undergoing some episode – this use should be carefully distinguished from ours.

A special case of CP, when $\Gamma = \emptyset$, is the claim

$\mathsf{CP}_\emptyset$ $Pr(\psi \mid \phi) = Pr(\phi \to \psi)$ for every rational ur-prior Pr.

Throughout this paper I shall adopt a shorthand of writing $\to$ to mean $\to_\emptyset$; informally I shall refer to $\to_\emptyset$ as the 'ur-conditional'. Most of the literature on probabilities of conditionals focus on principles which have the same form as $\mathsf{CP}_\emptyset$, so the primary focus of this paper is $\mathsf{CP}_\emptyset$ and not the seemingly stronger CP.[4] Once again, Pr ranges over rational initial probability functions and Γ over possible evidence.

A number of principles have been proposed in the literature relating probabilities of conditionals to conditional probabilities that look very much like $\mathsf{CP}_\emptyset$. In this tradition two have been widely discussed: Adams' thesis that the assertability of a conditional is the conditional probability of the consequent on the antecedent (see [1]) and Stalnaker's thesis that the probability of a conditional sentence is the conditional probability of the consequent on the antecedent (see Stalnaker [34].) Van Fraassen, McGee and others discuss versions of these theses with restrictions on what kinds of proposition can be substituted into the principle. These usually involve sentences in some way – a proposition obeys the relevant form of the principle only if it is expressed by a conditional that does not have conditionals embedded in it in certain ways. CP and $\mathsf{CP}_\emptyset$ are closely related to some of these theses, but there are some important differences.

$\mathsf{CP}_\emptyset$, for example, has the same form as Stalnaker's thesis except there is a restriction that Pr be an initial probability function. The restriction in $\mathsf{CP}_\emptyset$ weakens Stalnaker's thesis, telling us only about the conditional opinions of a person with no evidence whatsoever. To allow Pr to be substituted for a probability function representing the evidential probability or credences of someone with evidence would lead to a number of triviality results which we shall rehearse in the next section. $\mathsf{CP}_\emptyset$ on the other hand seems stronger than principles discussed by McGee and van Fraassen since there is no restriction to non-iterated conditionals.

As we have already mentioned there are a great number of impossibility results aimed at principles like $\mathsf{CP}_\emptyset$. My treatment of this literature shall therefore be far from comprehensive. The main line of defence against all of these kinds of arguments is the aforementioned technical argument that CP (and thus $\mathsf{CP}_\emptyset$) is consistent across a large class of ur-priors. However this response is not very illuminating, and it often pays to see how particular triviality arguments fail, and furthermore, it is important to show how the premises of these arguments are not philosophically motivated. The triviality results are often divided into two kinds – the 'dynamic' kind, which show the identity must fail under certain ways of updating probability functions, and 'static' results, which apply to a single probability function but which make assumptions about the conditional logic. In sections 1 and 2 I will address what I take to be the main versions of these two problems: Lewis's impossibility result [18], based on the assumption that the range of probabilities to which $\mathsf{CP}_\emptyset$ applies is closed under conditioning, and Stalnaker's impossibility result that appeals to his logic $\mathsf{C2}$. The upshot of these sections is that there is a very natural logic and semantics for conditionals

[4] However we show both principles to be consistent in the appendix.

which, for all these results show, are compatible with $\mathsf{CP}_\emptyset$ and CP. In section 3 the tenability result is proved – I provide a model of $\mathsf{CP}_\emptyset$ and CP within the semantic framework motivated in the earlier sections.

Before we start let me introduce some notation. I shall always use Pr to denote a rational ur-prior, i.e. a function representing a degree of confidence it would be rational to have if you have no evidence whatsoever. I shall use Cr to denote a rational credence function, i.e. a function which represents the degree of confidence it is rational for an agent to have on her evidence. Γ will be used to denote a set of propositions (or, equivalently, the conjunction of that set) which could possibly be some agent's total evidence and $\to$ shall always mean $\to_\emptyset$. When A is a proposition and Γ a set of propositions I shall write $Pr(A \mid \Gamma)$ for the conditional probability of A given the conjunction of the members of Γ. I shall often abuse notation by omitting set brackets – for example by writing $Pr(A \mid B)$ instead of $Pr(A \mid \{B\})$, $Pr(A \mid B, C)$ instead of $Pr(A \mid \{B, C\})$, and so on.

1 Stalnaker's thesis

'Stalnaker's thesis' is often understood as the claim that the credence one should have in an indicative sentence, 'if A then B', is the same as the conditional credence one should have in B given A, provided the latter is defined.

> Stalnaker's Thesis: For any rational credence function, Cr, $Cr(B \mid A) = Cr(A \to B)$ whenever $Cr(A) > 0$.

Since Stalnaker does not distinguish between the different connectives one might express with the indicative conditional, I shall temporarily relax reference to subscripts (we will worry about this soon enough.)

Stalnaker's thesis has been the subject a number of disproofs beginning with Lewis's [18]. But before I turn to these arguments I want to discuss another more basic problem with the thesis. Consideration of this problem should hopefully lead us to reconsider the data supposedly in support of Stalnaker's thesis, and to a better formulation of the thesis which explains the data but is not subject to disproof.

The problem is this. It is generally acknowledged among theorists working on conditionals that if indicative conditionals are to have truth conditions which are not truth functional, they must be context sensitive (Stalnaker [31], Gibbard [10], Nolan [24].) This is demonstrated, for instance, by the 'Sly Pete' example we outlined in (1) and (2) above. Which proposition is expressed by a conditional sentence depends on the context in which it is uttered.

The probability of a conditional sentence is therefore context sensitive in the following sense: the probability of a conditional sentence varies from context to context as the proposition it expresses varies. Thus, for example, the probability of the sentence 'I will live until I'm eighty five' is context sensitive – when it is uttered by someone with a healthy lifestyle they express a proposition with a greater probability than when it is uttered by an unhealthy person.

An indicative conditional, 'if A then B', can be context sensitive even when neither A nor B are context sensitive. In such cases neither the probability of 'A and B' nor the probability of A are context sensitive ('and' is not usually a source of context sensitivity.) So the number you get by dividing the probability of the proposition expressed by 'A and B' in a certain context by the probability of the proposition expressed by A in that context will be the same as the number obtained by performing the same calculation relative to any other context. The question 'what is the conditional probability of B given A?' has a quite definite, context invariant answer whereas the question 'what is the probability of if A then B' does not. We cannot therefore expect an arbitrary utterance of Stalnaker's thesis to be true: the number flanking one side of the identity will change from context to context whereas the number flanking the other side won't.

This raises a serious question about what claim Stalnaker's thesis is intended to state. Perhaps it states that the relevant conditional probability is the probability of the proposition expressed by the sentence 'if A then B' in some particular context: the context Stalnaker found himself in when he wrote [34], or at least the context we find ourselves in when we read [34]. On this interpretation we should read Stalnaker's thesis as making a very local claim about Stalnaker's own evidence when he wrote [34], namely that the probability of the proposition that is expressed by this use of 'if A then B' is the same as the probability of the proposition expressed by B given the proposition expressed by A in this context. I think it is clear that Stalnaker intended to capture something much more general than an autobiographical fact about himself so it is safe, I think, to disregard this possibility in what follows.

Stalnaker's thesis is context sensitive – it states that a context sensitive probability is the same as a context insensitive conditional probability – and provides no sensible context at which to evaluate it. Considerations such as these have led me to think that Stalnaker's thesis, at least as stated, cannot the correct theory of the probabilities of conditional sentences. However, it also seems obvious to me that our intuitions about the probabilities of conditionals in particular contexts, including those reported in (i)-(iv), are in *some way* connected to the corresponding conditional probability. The task, therefore, is to isolate this truth about the way we evaluate conditionals from Stalnaker's thesis.

Let me now formulate two revisions of Stalnaker's thesis in which the context sensitivity of conditional sentences is explicitly taken into account. I shall throughout assume that each context, c, provides a salient set of evidence, Γ, and as explained earlier I shall use $P \rightarrow_\Gamma Q$ for the proposition expressed in c by a conditional whose antecedent expresses P in c and whose consequent expresses Q in c. I shall also assume, following Lewis, that it is rational for an agent to be *Bayesian*. That is, that if Pr is a rational *initial* probability, and it is possible for Γ to be someone's total evidence, then $Pr(\cdot \mid \Gamma)$ is a rational credence function – i.e. a credence it is rational for some possible agent to have given their evidence. When we also assume that all rational people are Bayesian we can effectively eliminate talk of rational credences and just talk of functions of the form $Pr(\cdot \mid \Gamma)$ for rational ur-priors Pr and evidence Γ.

One principle that might do justice to the generality of Stalnaker's thesis says that Stalnaker's thesis holds of every rational credence function (i.e. every function of the form $Pr(\cdot \mid \Gamma)$) in every context. That is:

ST∀ $Pr(\psi \mid \Gamma \cup \{\phi\}) = Pr(\phi \rightarrow_\Sigma \psi \mid \Gamma)$ for every ur-prior Pr.

where Σ a set of propositions provided by any context you like and Γ any set of propositions which could be an agent's total evidence. In other words the connective $\rightarrow_\Sigma$ obeys the statement of Stalnaker's thesis for any rational credence, $Pr(\cdot \mid \Gamma)$.

It would of course be a miracle if Stalnaker's thesis were true in all contexts. Since the left hand side is sometimes context insensitive it could only be true if the propositions expressed by the right hand side in different contexts all happened to have the same probability for any agent. In my view the principle that best accounts for the data Stalnaker's thesis was supposed to account for says that when your total evidence is Γ (and, therefore, your credences are of the form $Pr(\cdot \mid \Gamma)$ for some ur-prior Pr) the probability of a conditional in a context where the evidence Γ is salient is the conditional probability of the consequent on the antecedent. In our formalism we get:

CP $Pr(\psi \mid \Gamma \cup \{\phi\}) = Pr(\phi \rightarrow_\Gamma \psi \mid \Gamma)$.

Note that CP is entailed by ST∀ (just set $\Sigma = \Gamma$) but is strictly weaker.

It is important to get clear what the data is that our answers to (i)-(iv) report. In the context in which those questions are asked a single piece of evidence is salient, the evidence of the subject, and we are evaluating these conditionals by her credences, which are, given our Bayesian assumptions, a rational ur-prior conditioned on this same piece of evidence. The answers reported to (i)-(iv) therefore confirm instances of CP. CP *only* predicts a relation between conditional credences and credences in conditionals when the agent's evidence and the contextually salient propositions are the same. Indeed, it is not hard to see why most of the data purportedly in support of Stalnaker's thesis are actually instances of the weaker principle CP: in most contexts the salient evidence *is* the utterer's evidence.

(i)-(iv) are also predicted by the stronger principle ST∀. In order to evaluate the extra strength of ST∀ one has to consider more contrived examples where the contextual evidence and the subject's evidence are different. It is thus hard to find counterexamples to ST∀ since most of our judgments about the probabilities of conditionals come about when the evidence that is salient at the time of the judgment is the evidence the determines what the relevant probabilities are. In order to see why ST∀ fails we must look to more theoretical arguments, such as the triviality results. The fact that it is hard to refute ST∀ with a direct counterexample points in favour of our current diagnosis; the cases where our judgments about probabilities of conditionals are at their firmest are the cases where ST∀ and CP coincide.

1.1 Lewis's triviality result

In [18] Lewis famously showed that *no* binary connective satisfies Stalnaker's thesis throughout a class of probability functions closed under conditioning unless the probability of a conditional is always the same as the probability of its consequent in that class. We say that a class is closed under conditioning if whenever Cr is in the class and $Cr(A) > 0$ then $Cr(\cdot \mid A)$ is also in the class. The result doesn't straightforwardly apply in this context since $A \to_\Gamma B$ is being treated as a *ternary* connective. However the result can be stated and discussed in this framework in terms of the binary connective $\to_\emptyset$, which we shorten to $\to$.[5]

Here we present Lewis's triviality result in a way that does not require anything as strong as Stalnaker's thesis. It requires only the substantially weaker thesis:

> You may be conditionally certain in B given A only if you are certain that if A then B.

While this principle sounds initially intuitive it suffers from the same defect as Stalnaker's thesis: its truth is context sensitive. For example, in the last section we argued that a principle like this can be false in a context in which all the salient evidence is tautologous (so that we may formalise 'if A then B' with $A \to B$) and where the agent in question has non-tautologous evidence. Lewis's result in effect establishes this fact directly.

Let us assume, then, that for any rational credence with $Cr(\phi) > 0$, if $Cr(\psi \mid \phi) = 1$ then $Cr(\phi \to \psi) = 1$. Reformulating this explicitly in terms of ur-priors gives us

1. If $Pr(\psi \mid \Gamma \cup \{\phi\}) = 1$ then $Pr(\phi \to \psi \mid \Gamma) = 1$ for every rational initial probability Pr and evidence Γ provided $Pr(\phi \mid \Gamma) > 0$.

Note that (1) entails the principle (*): if $Pr(\psi \mid \Gamma \cup \{\phi\}) = 0$ then $Pr(\phi \to \psi \mid \Gamma) = 0$ when $Pr(\phi \mid \Gamma) > 0$. The argument makes use of the uncontroversial inference from $\phi \to \neg\psi$ to $\neg(\phi \to \psi)$ when ϕ is (epistemically) possible.[6]

Now we make two simple observations. Trivially $Pr(B \mid A, B) = 1$ so it follows by (1) that $Pr(A \to B \mid B) = 1$. Equally trivially $Pr(B \mid A, \neg B) = 0$, so it follows that $Pr(A \to B \mid \neg B) = 0$ by (*). Therefore:

i. $Pr(A \to B) = Pr(A \to B \mid B)Pr(B) + Pr(A \to B \mid \neg B)Pr(\neg B)$ by probability theory.

ii. $Pr(A \to B \mid B)Pr(B) + Pr(A \to B \mid \neg B)Pr(\neg B) = 1.Pr(B) + 0.Pr(\neg B)$ by the two observations above.

[5] There is a further reason why we focus on $\to_\emptyset$ in what follows: in our eventual semantics $A \to_{\{E\}} B$ is equivalent to $A \wedge E \to B$ so we can in fact take the binary connective $\to_\emptyset$ as primitive.

[6] Suppose $Pr(\phi \mid \Gamma) > 0$ (so ϕ is epistemically possible). Then $Pr(\psi \mid \Gamma \cup \{\phi\}) = 0$, so $Pr(\neg\psi \mid \Gamma \cup \{\phi\}) = 1$. By 1 it follows that $Pr(\phi \to \neg\psi \mid \Gamma) = 1$. By the inference from $\phi \to \neg\psi$ to $\neg(\phi \to \psi)$, $Pr(\neg(\phi \to \psi) \mid \Gamma) = 1$ and finally $Pr(\phi \to \psi \mid \Gamma) = 0$ as required.

iii. $1.Pr(B) + 0.Pr(\neg B) = Pr(B)$

This result is disastrous: we have shown that if 1 is true the probability of any conditional is the same as the probability of its consequent. Any theory of probabilities of conditionals that commits us to 1 has to go.

What can we conclude about our two theories, $\mathsf{ST}\forall$ and CP? Both CP and $\mathsf{ST}\forall$ have special consequences for the connective $\rightarrow_\emptyset$, which for ease of use I shall name $\mathsf{ST}\forall_\emptyset$ and $\mathsf{CP}_\emptyset$:

$\mathsf{ST}\forall_\emptyset$ $Pr(\psi \mid \Gamma \cup \{\phi\}) = Pr(\phi \rightarrow \psi \mid \Gamma)$ for every ur-prior Pr.

$\mathsf{CP}_\emptyset$ $Pr(\psi \mid \phi) = Pr(\phi \rightarrow \psi)$ for every ur-prior Pr.

Note the difference between $\mathsf{ST}\forall_\emptyset$ and $\mathsf{CP}_\emptyset$. $\mathsf{ST}\forall_\emptyset$ applies to conditional ur-priors – to the credences of an informed agent – whereas $\mathsf{CP}_\emptyset$ only applies to ur-priors. This marks a crucial difference between the two theses. $\mathsf{CP}_\emptyset$ places a basic constraint on the rationality of your initial credences – provided you obey this constraint you will continue to be rational no matter what you update on. $\mathsf{ST}\forall_\emptyset$, on the other hand, tries to constrain your credences as you update by conditioning, and turns out to be inconsistent.[7]

Lewis's result is a problem for $\mathsf{ST}\forall_\emptyset$ but not for $\mathsf{CP}_\emptyset$. It is easy to see that $\mathsf{ST}\forall_\emptyset$ has 1 as a consequence whereas $\mathsf{CP}_\emptyset$ only has the weaker principle 2.

1. If $Pr(\psi \mid \Gamma \cup \{\phi\}) = 1$ then $Pr(\phi \rightarrow \psi \mid \Gamma) = 1$ for every rational initial probability Pr and evidence Γ.

2. If $Pr(\psi \mid \phi) = 1$ then $Pr(\phi \rightarrow \psi) = 1$ for every rational initial probability Pr.

To put it in Lewis's original language, the range of probabilities $\mathsf{CP}_\emptyset$ and 2 apply to are not closed under conditioning. They apply only to initial evidential probability functions representing permissible evidential probabilities of agents with no evidence at all. If the initial probability distributions were not regular then they would represent the possibility of an agent who has been able to rule out some contingent or empirical hypothesis *a priori* without having received any evidence. To rule out an emprical hypothesis without evidence is irrational. No set of regular probability functions is closed under conditioning since the result of conditioning a probability function on a contingent proposition is never regular.

[7]Unfortunately it appears to be quite easy to mistake a principle formulated in terms of ur-priors for one formulated in terms of a rational agent's current credences. For example, a common mistake regarding the Principal Principle, according to Meacham, 'replaces the reasonable initial credence function that appears in Lewis' original formulation with a subject's current credence function' ([23]) The mistake is just as disastrous in this case and leads to inconsistencies. The Principal Principle, like $\mathsf{CP}_\emptyset$, are in some sense *a priori* constraints and should not rule out an agents recieving this or that piece of empirical evidence.

1.2 Conditional proof

According to the theory we endorse certain modes of inference are acceptable for people who have no evidence which are unacceptable for people who have particular evidence. This is seen, for example, from the fact that we endorse 2 but not 1 (and $\mathsf{CP}_\emptyset$ but not $\mathsf{ST}\forall_\emptyset$.) When one has no evidence at all one should evaluate the probability of $A \rightarrow_\emptyset B$ as the same as the probability of B on A, but once further evidence E is in this connection may not hold (although a similar connection between the conditional probability and the probability of $A \rightarrow_{\{E\}} B$ will hold.) Other writers have suggested that to correctly capture conditional reasoning we must keep track of which background suppositions or which background evidence is in place (see for example McGee [22] and Humberstone [14] §7.16.) For example, conditional logics typically distinguish between conditional proof with side premises and conditional proof without, usually treating only the latter as valid. In this section I show how this phenomenon is closely connected with the conditional epistemology defended here.

Let's start by highlighting an analogy between the two probabilistic principles we have been discussing, namely:

1. If $Pr(\psi \mid \Gamma \cup \{\phi\}) = 1$ then $Pr(\phi \rightarrow \psi \mid \Gamma) = 1$ for every rational initial probability Pr and evidence Γ.

2. If $Pr(\psi \mid \phi) = 1$ then $Pr(\phi \rightarrow \psi) = 1$ for every rational initial probability Pr.

and two logical principles which state versions of conditional proof, with and without side premises:

1′. If $\Gamma, \phi \vdash \psi$ then $\Gamma \vdash \phi \rightarrow \psi$

2′. If $\phi \vdash \psi$ then $\vdash \phi \rightarrow \psi$

1′ says that if you can validly infer ψ from $\Gamma \cup \{\phi\}$ then you can validly infer $\phi \rightarrow \psi$ from Γ, whereas 1 says that if you're conditionally certain in ψ given ϕ against background evidence Γ, you should be certain in $\phi \rightarrow \psi$ with background Γ. 2′ says that if you can validly infer ψ from ϕ you can always validly infer $\phi \rightarrow \psi$ from no assumptions whereas 2 says that if you are conditionally certain in ψ given ϕ, when you have no evidence, you should be certain in $\phi \rightarrow \psi$. 2 and 2′ are special cases of 1 and 1′ respectively (let $\Gamma = \emptyset$.) It is very natural to think that these two pairs of principles are closely related. For example, if 1 is true then 1′ represents a good form of argument in the following sense: if your initial credences recommend being fully confident in ψ conditional on the premises $\Gamma \cup \{\phi\}$ then they will recommend being fully confident in $\phi \rightarrow \psi$ conditional on Γ. Similarly if 2 is true then 2′ also represents a good form of inference in the sense that if you're initial credences recommend being conditionally certain in ψ given ϕ they will recommend being certain that $\phi \rightarrow \psi$ simpliciter.

What about the converse connection? For example, if 2′ is a valid form of inference must 2 be true? The matter is less straightforward, but I think an argument can be made. Suppose that Pr is a rational initial probability and $Pr(\psi \mid \phi) = 1$. It follows by calculation that $Pr(\phi \wedge \neg\psi) = 0$. Since Pr is regular, in the sense that it assigns every possibility a non-zero value, it follows that $\phi \wedge \neg\psi$ is not a genuine possibility, i.e. $\phi \vdash \psi$. By $2'$ it follows that $\phi \to \psi$ is valid, so $Pr(\phi \to \psi) = 1$.[8] A similar argument could be made for 1 on the basis of $1'$.

There is an interesting difference between $1'$ and $2'$, however, which sheds some light on the difference between 1 and 2. $2'$ is valid and $1'$ is not. The case for the validity of $2'$ is both intuitive and evidenced by the fact that it is validated on all the standard possible world semantics for conditionals. Putting aside irrelevant differences between the possible world theories, they all say, give or take, that a conditional is true at a world if the consequent is true at the closest world (or worlds) to it at which the antecedent is true. So if every ϕ world is a ψ world then, for any world, the closest ϕ world to it is a ψ world.

$1'$, however, is invalid. The case for this is again supported by intuition and the possible world semantics. We'll begin with the latter. Note that in any model, ϕ and ψ entail ϕ (i.e. every ϕ and ψ world is a ϕ world) so if 1 were a sound rule we should have that ψ entails $\phi \to \psi$. This is not so; there could be a ψ world which is not a ϕ world, and where the closest ϕ world is not ψ either. Intuitive counter examples to 1 abound. For example note that:

> I will wake up at 7am on Saturday, I drink a bottle of vodka on Friday night; therefore I will wake up at 7am on Saturday

is a valid inference, whereas the following is patently false:

> I will wake up at 7am on Saturday; therefore if I drink a bottle of vodka on Friday night I will wake up at 7am on Saturday.

I may well wake up at 7am on Saturday, but not if I drink a bottle of vodka the night before. The presence of side premises in $1'$ therefore marks an important difference between $1'$ and $2'$.

There is therefore a principled reason why we should accept 2 and reject 1. 2, unlike 1, is in fact *guaranteed* by general norms governing the way in which our conditional confidence is regulated by validity: as we have argued $2'$ is valid, and that if $2'$ is valid then 2 is true. On the other hand 1 appears to inherit the counterexamples to $1'$: I should not be fully confident, even conditional on the supposition that I wake up at 7 on Saturday, that I'll wake up at 7 on Saturday if I drink a bottle of vodka the night before.[9] This directly contradicts 1 since my conditional confidence that I'll wake up at 7 given I'll wake up at 7 and I drink a bottle vodka is, of course, 1.

[8] The potentially controversial step is where we inferred that $\phi \vdash \psi$ from the fact that $\phi \wedge \neg\psi$ is false in all possibilities. I have tacitly assumed that a rational initial probability is one which assigns every *logical* possibility a non-zero value. If this is assumed is then $2'$ entails 2 (and similarly $1'$ entails 1.)

[9] Although I concede, and predict, that this judgment is context sensitive. In a context in which I somehow have knowledge that I'll wake up at 7 I am much less inclined to assign a low conditional probability to this sentence.

However, notice that there is a second reading of the above counterexample which is slightly harder to get, but which can be accounted for by taking the context sensitivity of conditional statements into account. Suppose, now, that the fact that I will wake up at 7am on Saturday is really *part of my evidence* – let's suppose some completely reliable oracle has told me that I'd wake up at 7. Now it seems OK to say 'look, the oracle told me that I'd wake up at 7. So, whatever happens tonight, even if I drink a bottle of vodka, I'll wake up at 7.' If we add the proposition that I wake up at 7 on Saturday to the evidence we evaluate the conditional with respect to we can infer that if I drink a bottle of vodka on Friday, I'll wake up at 7 tomorrow, in the contextually salient sense of 'if'. In our current notation we can capture this form of reasoning by the following version of conditional proof:

$3'$ If $\Gamma, \phi \vdash \psi$ then $\Gamma \vdash \phi \to_{\Gamma} \psi$

This version of conditional proof is, as with $1'$ $2'$, closely connected to a probabilistic principle, namely

3. If $Pr(\psi \mid \Gamma \cup \{\phi\}) = 1$ then $Pr(\phi \to_{\Gamma} \psi \mid \Gamma) = 1$ for every rational initial probability Pr.

3 is a consequence of CP.

Conditional proof is also closely related to the infamous 'or-to-if' argument. Consider the following instance of $3'$

$\phi \vee \psi, \neg\phi \vdash \psi$

Therefore $\phi \vee \psi \vdash \neg\phi \to_{\phi\vee\psi} \psi$

The conclusion sequent is an instance of 'or-to-if' reasoning, and is often motivated by examples such as the following: either the butler did it or the gardener did; therefore if the butler did not do it the gardener did. Traditional formulations of the 'or-to-if' argument employ a binary connective, which, on the basis of this inference, can be shown to be equivalent to the material conditional (assuming also modus ponens.) However, if we accept that the conclusion of this argument is context sensitive in the way I have suggested, the current formulation seems to fare better, and indeed does not collapse the conditional into material implication.[10]

In short, then, we see that $\mathsf{CP}_{\emptyset}$ evades Lewis's results due to its restriction to ur-priors. However, we have argued that this restriction is not *ad hoc*: it is part of a general theory (CP) for evaluating probabilities of context sensitive conditional statements, and is also motivated by accepted principles of conditional logic in a way in which the unrestricted version ($\mathsf{ST}\forall_{\emptyset}$) isn't.

Unfortunately, however, there are other triviality results which do not rely on 1. These use the *logic* of conditionals to cause problems, and it is to these I now turn.

[10]For the origin of this sort of response to the 'or-to-if' argument see Stalnaker [31], who endorses 'or-to-if' speechs as pragmatically justified, albeit formally invalid. Here we give 'or-to-if' its due, as a formally valid inference by explicitly treating the conditional as a three place relation.

2 Stalnaker's impossibility result

In [32] Stalnaker, rejecting the thesis he once held, provided another triviality result which did not rely on Lewis's closure assumption. This is a 'static' triviality result: unlike Lewis's result, which required Stalnaker's thesis to continue to hold under various ways of updating probabilities, Stalnaker's argument shows that no single probability function can satisfy the thesis. Unlike Lewis's result, this kind of argument seems to be directly applicable to $\mathsf{CP}_\emptyset$. However this argument also relies on contentious principles of conditional logic.

Stalnaker's argument was later refined by Hájek and Hall in [12] who isolated principles of Stalnaker's $\mathsf{C2}$ that are sufficient for the proof. One of these principles represents a weakening of transitivity, $((\phi \to \psi) \wedge (\psi \to \phi) \wedge (\psi \to \chi)) \supset (\phi \to \chi)$, which is often referred to as CSO. Here $\to$ represents the indicative conditional and $\supset$ the material conditional.

Indeed, CSO appears to be the primary culprit for this result, which can be seen from inspecting the assumptions of Hájek and Hall's result.

Theorem 2.1. *Suppose that* $\mathsf{CP}_\emptyset$ *holds and that the following principles are valid*

MP $\phi, \phi \to \psi \vdash \psi$

CC $\phi \to \psi, \phi \to \chi \vdash \phi \to (\psi \wedge \chi)$

CSO $((\phi \to \psi) \wedge (\psi \to \phi) \wedge (\psi \to \chi)) \supset (\phi \to \chi)$

Then no more than two disjoint consistent propositions have positive probability for any rational ur-prior.

For the argument see [12]. I take it that modus ponens is non-negotiable[11] and that CC is equally self-evident. So the upshot of theorem 2.1 is clear: either $\mathsf{CP}_\emptyset$ goes or CSO goes. The case against $\mathsf{CP}_\emptyset$ is therefore only as good as the case for weakened transitivity.

What does weakened transitivity have to recommend itself? It is a bit hard to say what its intuitive rationale is without appealing to the transitivity of $\to$. (That CSO is a weakening of the transitivity of $\to$ can be seen by deleting $\psi \to \phi$ from the antecedent.) This would be a bad basis on which to justify weakened transitivity since transitivity is itself considered to be invalid by many, including Stalnaker himself (see [30].) For example, suppose we know that an experiment involving a match has taken place, although we do not know whether the match was struck. We agree that if the match was struck then it lit. Furthermore, we clearly agree that if the match was soaked in water and struck, it was struck, since the antecedent entails the consequent. But transitivity would allow us to infer that if the match was soaked in water and struck then it lit. In other words transitivity allows us to commit the error of strengthening the antecedent.[12]

[11] Although for scepticism about modus ponens see McGee [21].

[12] Stalnaker also provides direct counterexamples to transitivity which do not involve antecedent strengthening.

In the following subsections I shall make a case for the view that CSO is not valid on the basis of independent, non-probabilistic considerations. I shall then argue that, while CSO is validated in many standard possible world semantics (see Lewis [17], Stalnaker [30], Pollock [26]), which appeal to what happens at the closest antecedent-worlds to the world of evaluation, one can construct a slightly different possible worlds semantics that can accommodate the counterexamples without detracting much from the theoretical simplicity of these accounts.

We should think of the following counterexamples as *prima facie* cases against CSO and related principles. The purpose of this paper, however, is to defend the tenability of CP, not to conclusively refute CSO, and I shall accordingly not spend much time considering potential responses one could make to these counterexamples. My purpose is to rather motivate alternatives to Stalnaker's semantics.

2.1 Counterexamples: subjunctives

It is worth noting that CSO is one of the more controversial commitments of the influential Lewis-Stalnaker account conditionals. Indeed, there is a small but substantial body of underdiscussed papers in which counterexamples to this principle have been suggested (see [35], Stalnaker [REF], [20], Tooley [36] and Ahmed [2].[13]) In addition to these counterexamples there are general theoretical reasons for dropping CSO (see, for example, analyses of conditionals which take the notion of causal independence seriously: Kvart [16], Martensson [20], Schaffer [28], Edgington [9].)

In this section I'll present my own counterexample to the subjunctive version of CSO. In the next section I'll consider the indicative version. In both cases, I hope, the countexamples ought to make it clear how Stalnaker's semantics fails and how it can be modified.

In order to state his formal semantics (which we shall describe shortly) Stalnaker introduces a selection function, $f(A,x)$, mapping a proposition, A, and a world, x, to another world. Neutrally speaking, $f(A,x)$ represents the world that would have obtained (instead of x) had A obtained. Somewhat less neutrally, Stalnaker identifies $f(A,x)$ with the closest world to x at which A obtains, on a certain technical understanding of 'close'. On this second interpretation it follows that if $f(A,x)$ is a B-world (a world in which B obtains) and $f(B,x)$ is an A-world then $f(A,x) = f(B,x)$ – if the closest A-world is a B-world and the closest B-world is an A-world then the closest A-world *is* the closest B-world. It is exactly this property of Stalnaker's selection function that ensures the validity of CSO in the formal semantics.

But now consider the set up in figure 1. Here is how things would have gone if A had toppled: A would have rolled down the slope, up the opposite slope toppling B from its perch. Here is how things would have gone if B had

[13]It's worth also noting Pollock [27] p254 in this regard. Pollock presents a counterexample to a principle that is a consequence of CSO and conditional excluded middle.

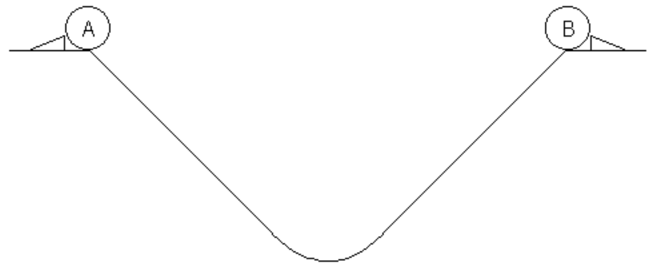

Figure 1: Two balls A and B balanced on opposing slopes.

toppled: B would have rolled down the slope, up the opposite slope toppling A from its perch. These are clearly two different scenarios. In actual fact neither A nor B topple. If @ is the actual world and $f(X, @)$ the way things would have gone had X obtained then, although the world that would have obtained had A toppled is a world where B topples, and *vice versa*, the way things would have gone had A toppled is clearly not the same as the way things would have gone had B toppled.

We can turn this into a counterexample to CSO as follows. Suppose that A and B are positioned on two pressure plates. If the pressure is released on A's plate then the mechanism on B's plate will be immediately deactivated and a green light will come on and remain on. If the pressure on B's plate is released then the mechanism on A's plate will immediately deactivated and a red light will come on and remain on. It is impossible for both the green and red light to be on at once since one of the two mechanisms will be deactivated before the other, depending on which ball topples first. Now suppose we can see that neither ball has toppled. It seems to me that it would be perfectly felicitous to go on and assert the following:

1. If A were to topple, B would topple.
2. If B were to topple, A would topple.
3. If A were to topple the green light would come on.
4. If B were to topple the red light would come on.

Yet according to CSO (1-4) are inconsistent. From 1-3 we can infer that if B were to topple the green, and not the red, light would come on, contradicting 4.[14]

[14]There is a question, which I cannot fully address here, as to what role tense is playing in these examples. This issue could be circumvented altogether if one could replace 1 with 'if A were to topple at some point, B would topple at some point', and make similar substitutions in 2-4, without disturbing the truth values of 1-4. It is unclear to me whether the resulting fours sentences would all be true.

Let me first concede that it is possible to prime the audience to read 3 (or, indeed, 4) in a way that makes it less acceptable. I might say to you: 'look, when A topples because it's hit by B the green light will not come on because its mechanism has been deactivated. So if A were to topple, the green light might not come on (because B toppled first) so we should not accept 3.' This argument relies on a controversial connection between 'might' and 'would' counterfactuals, but an assumption I am willing to grant for the sake of argument.

The kind of reasoning described above is sometimes called 'backtracking'. I am inclined to agree with Lewis [19], and many after him, that the contexts in which backtracking is legitimate are somewhat special contexts. As Lewis puts it: 'A counterfactual saying that the past would be different if the present were somehow different may come out true under the special resolution of its vagueness, but false under the standard resolution.' The same distinction applies to the 'might' counterfactual we used in the above argument.[15]

It must be stressed, however, that our case against CSO does not depend on the possibility of contexts in which backtracking is legitimate. We only need 1-4 to be simultaneously true in one context to complete our case against CSO, which deems them jointly inconsistent. It does not matter if there are also contexts in which some or all of 1-4 are false.

2.2 Counterexamples: indicatives

To what extent do these counterexamples generalise to indicative conditionals? Given the assumption that indicative and subjunctive conditionals have the same logic, it is clear that the above counterexample has a direct bearing on indicatives. For what it is worth, my view is that this assumption is entirely reasonable, even given the differences between the two classes. However, I don't want to rest my case on this assumption so I shall consider the indicative case separately here.

There is one straightforward sense in which these counterexamples generalise. People have often noted that if I say now 'if A topples, B will topple' you can later report my speech by saying '[XXXX] said that if A had toppled, B would have toppled' (see Dudman [8].) Thus subjunctive conditionals often express in the past tense what a 'forward looking' indicative conditional (i.e. a conditional containing the future modal 'will' in the consequent) can be used to express at an earlier time. Thus if in figure 1 what I say now with the sentences 1-4 can be said at an earlier time by uttering appropriate indicative conditionals then the counterexample clearly generalises to the indicative case.

Perhaps all this shows is that the distinction between indicative and subjunctive doesn't cut at the semantic joints. Another question that is worth investigating is whether these counterexamples generalise to bare indicatives – indicatives with no visible modal in the consequent ('will', 'must', 'should' et cetera.) Suppose that we know that a system was set up as in figure 1 an hour

[15] Indeed, according to the logics containing conditional excluded middle there is little difference between 'would' and 'might' counterfactuals assuming, as we did above, that they are duals of one another.

ago, but we don't know how the state of the system has changed since. Then it seems that the following four conditionals may, for all I know, be all true:

1. If A toppled, B toppled.
2. If B toppled, A toppled.
3. If A toppled the green light came on.
4. If B toppled the red light came on.

However, according to CSO, these are inconsistent and therefore cannot, for all I know, be simultaneously true.

The consistency of 1-4 is one thing, whether they are ever jointly assertable is another. Here is an argument that 1-4 are jointly assertable only if you know their antecedents to be false:

> I cannot assert 3 unless I can rule out the possibility that A toppled and the green light didn't come on.[16] But I know that any world where B toppled initially, later colliding with A, is a world where A toppled and the green light doesn't come on, so I can rule out that B toppled first. By symmetrical reasoning I know that I can assert 4 only if I can rule out the possibility that A toppled first. So I may assert 3 and 4 only if I know that neither A nor B toppled.

However if the antecedents of 1-4 are known to be false this ruins the example as indicatives whose antecedents are known to be false are not usually assertable.[17]

This problem does not generalise to subjunctives: it is perfectly fine to assert, for example, that if A had toppled B would have toppled, even when you know that A hasn't toppled. The most this proves is that the pragmatics of indicative and subjunctive conditionals are different. Indeed I want to go further and suggest that although the indicative form of CSO is not semantically valid it states an important constraint on the kind of inferences that are pragmatically appropriate. Say that an inference from sentences $A_1 \ldots A_n$ to the sentence B is pragmatically appropriate iff, whatever the context of utterance, the conjunction of $A_1 \ldots A_n$ is less probable for the agent of that context than B is in that context. It follows that if the inference from $A_1 \ldots A_n$ to B is pragmatically appropriate then $A_1 \ldots A_n, \neg B$ cannot all have probability 1 in a context of utterance. Assuming CP it is possible to show that CSO is pragmatically appropriate.

[16] I cannot assert that if A toppled then B toppled unless I know it. So every epistemically possible world is a world where B toppled if A did, and thus, there are no epistemically possible worlds where A toppled and B didn't.

[17] Sometimes it is possible to read conditionals whose antecedents are known to be false as vacuously true. For example, if I know that A won't topple I can assert: if A topples then I'm a monkey's uncle. In contexts where the conditionals are vacuously true, however, we cannot infer 'it's not the case that if B toppled the green light came on' from 'if B toppled the red light came on' so the counterexample to CSO fails.

2.3 Conditional Logic and Semantics

I do not want to claim that these examples spell the end for Stalnaker's theory; there is still room for manoeuvre and considerations to be made in its favour. My aim, more modestly, is to dislodge the idea that CSO is an incontrovertible principle of conditional logic.

I would conjecture that CSO is not accorded this status on the basis of direct intuitions about its validity, but rather on the basis of more holistic theoretical considerations. One point in favour of the principles is that it is validated by a well known semantics that promises to give a general theory of conditionals. If we are to reject this theory we had better offer an alternative. In the following sections I'll suggest a couple of ways to develop a possible world semantics, formally similar to Stalnaker's, which does not validate the problematic principles.

We shall work with a toy modal propositional language, $\mathcal{L}$, consisting of the usual truth functional connectives, $\neg$ and $\supset$ from which the other truth functional connectives are definable, and a special binary modal connective representing the ur-conditional, $\rightarrow$. A *frame* for $\mathcal{L}$ is a pair $\langle W, f\rangle$ where W is a set of worlds and $f : \mathcal{P}(W) \times W \rightarrow \mathcal{P}(W)$ – f is called the 'ur-selection function'.[18] A model is a pair $\langle \mathcal{F}, [\![\cdot]\!]\rangle$ where $\mathcal{F}$ is a frame and $[\![\cdot]\!]$ maps propositional letters to subsets of W. $[\![\cdot]\!]$ extends to a function from the rest of $\mathcal{L}$ to $\mathcal{P}(W)$ as follows:

- $[\![\neg\phi]\!] = W \setminus [\![\phi]\!]$
- $[\![\phi \supset \psi]\!] = (W \setminus [\![\phi]\!]) \cup [\![\psi]\!]$
- $[\![\phi \rightarrow \psi]\!] = \{w \mid f([\![\phi]\!], w) \subseteq [\![\psi]\!]\}$

Logics based on this semantics can described quite neatly. Here are some principles which we shall be focusing on.

RCN if $\vdash \psi$ then $\vdash \phi \rightarrow \psi$

RCEA if $\vdash \phi \equiv \psi$ then $\vdash (\phi \rightarrow \chi) \equiv (\psi \rightarrow \chi)$

CK $(\phi \rightarrow (\psi \supset \chi)) \supset ((\phi \rightarrow \psi) \supset (\phi \rightarrow \chi))$

ID $\phi \rightarrow \phi$

MP $(\phi \rightarrow \psi) \supset (\phi \supset \psi)$

C1 $(\phi \rightarrow \psi) \supset ((\psi \rightarrow \bot) \supset (\phi \rightarrow \bot))$

CEM $(\phi \rightarrow \psi) \vee (\phi \rightarrow \neg\psi)$

CA $((\phi \rightarrow \chi) \wedge (\psi \rightarrow \chi)) \supset (\phi \vee \psi \rightarrow \chi)$

RCA $(\phi \vee \psi \rightarrow \chi) \supset ((\phi \rightarrow \chi) \vee (\psi \rightarrow \chi))$

[18] See Chellas [6]. One can define the selection function for other conditional connectives of the form $\rightarrow_\Gamma$ as $f_\Gamma(A, x) = f(\bigcap \Gamma \cap A, x)$.

VLAS $(\phi \to \psi) \supset ((\phi \to \chi) \supset (\phi \wedge \psi \to \chi))$

LT $(\phi \to \psi) \supset ((\phi \wedge \psi \to \chi) \supset (\phi \to \chi))$

CSO $((\phi \to \psi) \wedge (\psi \to \phi) \wedge (\phi \to \chi)) \supset (\psi \to \chi)$

The logic CK denotes the logic consisting of the rules RCN, RCEA and the axiom CK. CK is validated in every frame, and is analagous to the smallest normal modal logic K.[19] Indeed, we can consider it as a multi-modal logic in which $\phi \to$ is a normal modal operator for each substitution of ϕ.

Stalnaker's logic C2 includes all of the above principles, and Lewis all except for CEM. Stalnaker guarantees this by placing further constraints on the function f, namely:

F1. $w \in f(A, w)$ whenever $w \in A$

F2. $f(A, w) \subseteq A$

F3. $f(A, w) \subseteq \{x\}$ for some x

F4. if $f(A, w) \subseteq B$ and $f(B, w) \subseteq A$ then $f(A, w) = f(B, w)$

F1 ensures the validity of MP, F2 ensures ID, F3 ensures CEM and F4 ensures the weakened transitivity axiom, CSO. F4 also guarantees that the weaker principles CA, RCA, LT and VLAS are valid.[20] In the presence of F3 we can modify the semantics in to Stalnaker's original form so that f maps us from a world, w, and a set of worlds, A, to a single possible world (namely x if $f(A, w) = \{x\}$ in the general semantics) or the unique impossible world λ (if $f(A, w) = \emptyset$) in the general semantics) at which every sentence is stipulated to be true.

In the next sections we shall focus on two problems with Stalnaker's semantics. Both, I argue, require relaxing the constraint F4 and giving up CSO. The first problem is that the relevant aspects of similarity rarely ever pick out a *unique* world maximally similar to the actual world. While Lewis and others have given up constraint F3 in order to accommodate this, this strategy has recently been shown to give the wrong predictions in non-deterministic worlds. I shall argue that the best way to accommodate these issues is to keep F3 and relax F4 instead. The second problem is that Stalnaker's semantics validates CSO and therefore cannot accommodate the putative counterexamples we have described for subjunctives and indicatives.

---

[19] As with K we use the same name for the logic and its characteristic axiom.

[20] Stalnaker also stipulates that $f(A, w) = \emptyset$ only if $A = \emptyset$. This is there for a technical reason: it ensures that the notion of epistemic possibility can be modelled by a universal accessibility relation. However, since any reasonable notion of epistemic necessity will not iterate in an S5 fashion this constraint does not seem to be particularly motivated and I shall ignore it in what follows.

2.3.1 Non-determinism

For us it is the last restriction, F4, that is responsible for the problematic principles we have been considering. On Stalnaker's preferred reading $f(A,x)$ denotes the closest world to x at which A is true: $f(A,x)$ is a a world in which A is true but which differs minimally from x in respects of law and particular fact. F4 states that if the closest A world is a B world and if the closest B world is an A world then the closest A world *is* the closest B world. This principle ought to be obviously true if we are thinking of $f(A,x)$ as denoting the closest A world to x, where 'closeness' here denotes an independent ordering that does not depend on the proposition A being evaluated.

However, there is a well known objection to Stalnaker's semantics. Why should there be a *unique* closest world at which A is true? Surely there could be multiple worlds equally close with respect to the relevant factors? For this reason Lewis drops condition F3 in the general semantics and allows $f(A,x)$ to sometimes denote a set with two or more members.[21] I shall focus on the case of counterfactuals in what follows, since that is how the debate has been framed (although, as before, the discussion generalises to indicatives.)

Lewis's response involves relinquishing the principle CEM, a principle I take to be highly desirable (although this is not the place to defend it.) This is problematic for those moved by the kinds of epistemological considerations outlined here since CEM is probabilistically valid given $\mathsf{CP}_\emptyset$.[22]

There are, however, more pressing problems with Lewis's response. According to Lewis, a conditional is false if *any* of the closest antecedent worlds is not a consequent world. However a number of authors (Hawthorne [13], Hájek [11]) have noted that if the laws are chancy, as predicted by quantum mechanics say, there will be, among the closest worlds, all kinds of wild possibilities. For example, worlds where a plate is dropped and, instead of hitting the floor, floats upwards, represent no more of a departure from actual laws than a world in which it falls and hits the ground. The claim that the plate would break if dropped seems to be highly probable, even granting the laws are this way, because although not impossible it is very improbable that the plate would float upwards. The claim that the plate would break if dropped seems to be at least as probable as the claim that the plate will break, asserted after the plate is dropped and before it breaks (or floats upwards!) Lewis's analysis, however, predicts that this conditional is outright false: if the plate floats upward in any of the closest worlds then the counterfactual is false. If I am certain that among the closest worlds there are worlds where the plate floats upwards, as I should be if I believe quantum mechanics, I should reject the claim that the plate would break if dropped – I should be certain it is false.

There is, however, another way to respond to the problem for Stalnaker's

[21] This modification does not address the problem of there being no closest world due to there being an infinite succession of closer and closer worlds. Lewis addresses this in other papers.

[22] For example, if ϕ is consistent, $Pr((\phi \to \psi) \vee (\phi \to \neg\psi)) = Pr(\phi \to \psi) + Pr(\phi \to \neg\psi) - Pr((\phi \to \psi) \wedge (\phi \to \neg\psi)) = Pr(\psi \mid \phi) + Pr(\neg\psi \mid \phi) - 0 = 1$ for every rational initial prior Pr, and therefore $= 1$ for every informed prior with $Cr(\phi) > 0$ too.

theory: instead of giving up the constraint F3 relinquish F4.[23] We agree with Lewis that, in the relevant sense of closeness (keeping in accordance with particular matter of fact, respecting natural laws, and so on) there will typically be multiple closest worlds to the actual world at which a proposition is true. However, we agree with Stalnaker that there's a particular way things would have turned out if A had obtained. In order to evaluate a conditional, $A \to B$, in such a case, we look to the world that *would* have obtained if A had obtained, and this will be a single member of this set. (It is particularly hard to deny conditional excluded middle for indicative conditionals. Either the coin landed heads if it was flipped, or it didn't; there just doesn't seem to be room for a third option.)

On this proposal we therefore sever the straightforward connection between the function $f(A, x)$, which is used in the formal semantics, and counterfactual similarity. $f(A, x)$ represents the world that would have obtained instead of x if A had obtained. If there is exactly one closest world, w, then the way things would have gone had A obtained is just w. However, when there are multiple worlds that are equally close in the relevant respects (for example, if the laws are chancy) then the way things would have gone had A obtained is not determined by the closeness facts. If, for example, x is a chancy world, then there will be an element of randomness – the way things would have gone will certainly be one of the closest worlds, but due to the element of chance this cannot be deterministically calculated from non-counterfactual facts about what's happened at x and closeness facts. Thus, according to this alternative, the world $f(A, x)$ which is selected is not determined by the closeness facts – it is primitively counterfactual and it is an irreducibly chancy matter which, among the closest worlds, this world is.

It is clear why this does better than Lewis's theory. On this semantics the claim that the plate would break if dropped is highly probable, and is as assertable as anything is in a quantum world, whereas for Lewis it is false with probability 1. Even for Stalnaker, it seems, the conditional is indeterminate with probability one, thus also unassertable (assuming that it is not permissible to assert something you're certain is indeterminate.)

Finally, this semantics also allows us to see why F4 is refuted and why CSO is invalid. Suppose that the closest A worlds are B worlds and that the closest B worlds are A worlds. It follows by simple facts about closesness that the closest A worlds *are* the closest B worlds (this is what motivates principle F4 in the Lewis/Stalnaker semantics.) Call the set of closest A/B worlds X. Now note that the member of X that would have obtained if A had obtained is non-deterministically selected from X and thus might not be the same as the member of X that would have obtained had B obtained. Thus $f(A, x)$ need not be the same as $f(B, x)$ in cases like this, even though the world that would have

[23] The idea behind this semantics is due to Moritz Schulz [REF]. Schulz's theory places a lot of emphasis on the theory of arbitrary reference and epsilon terms which I am not following closely here. However, the basic idea – that $f(A, x)$ represents a randomly selected world from the closest A worlds to x – is the same. Schulz, however, prefers to add the constraint F4 by brute force, even though it does not appear to be motivated by his semantics.

obtained if A had obtained, $f(A, x)$, is a B world, since by hypothesis all the closest A worlds are B worlds, and also the world that would have obtained if B had obtained, $f(B, x)$, is an A world, since we know the closest B worlds are A worlds. In other words, $f(A, x) \in B$ and $f(B, x) \in A$ even though $f(A, x)$ may not be identical to $f(B, x)$.

The proposal extends to indicatives in a similar way. $f(A, x)$ represents the way the world will turn out if A happens. Even on an epistemic interpretation of 'closeness' there will be cases where there are multiple equally close worlds. In these cases the world that will obtain if A obtains is one of these worlds, but not any particular one with evidential probability 1.

2.3.2 Determinism

The counterexamples we have discussed (and those mentioned by Tichý, Pollock, Martensson, Tooley and Ahmed) all appear to be compatible with determinism so it is unclear whether the suggestion above is the best way to accommodate these examples. A number of people have suggested that we instead relativise the notion of closeness to the antecedent of the conditional (see Ahmed [2] and Noordhof [25].) According to this proposal a conditional $A \to B$ is true only if B is true at the A-worlds which are closest according to an ordering determined in part by A.

As an example of an antecedent relative theory Ahmed [2] points us to the theory of Edgington [9] which is tied to a tradition of taking causation as primitive in the analysis of counterfactuals (see also Kvart [16], Martensson [20], Schaffer [28], [9] all of whom seem to endorse antecedent relativity of some sort or other.) In listing what counts for closeness Schaffer, who is quite up front about the invalidity of CSO, writes that among other things we should try to "maximize the region of perfect match, from those regions causally independent of whether or not the antecedent obtains."[24]

In fact the thought that the similarity relation depends on the antecedent is quite pervasive in the literature on counterfactuals for the reason that on most accounts similarity depends on the time of the antecedent. On such accounts a subjunctive conditional $A \to C$ is true when and only when C is true at the A-worlds that are like the actual world in matters of particular fact up to the antecedent time and continue according to the causal laws after the antecedent time. According to Bennett this 'general idea has been accepted by all analysts of subjunctive conditionals in the 'worlds' tradition, and by most others as well.' [4] p198.[25]

The consensus seems to be that differences of particular fact after the antecedent time count for very little compared to differences before the antecedent

[24]See also Cross [7] for a discussion of the formal semantics (albeit from an unsympathetic perspective.)

[25]Whether this idea makes its way into the semantics endorsed by these theorists is another matter. For Lewis, for example, the dependence on an antecedent time is provided by the pragmatics. See Bennett [4] §118 for discussion.

time, when it comes to counterfactual similarity. This fact is nicely demonstrated by Fine's example:

> If Nixon had pressed the button, then there would have been a nuclear holocaust.

Worlds where there has been a nuclear holocaust are dramatically different from our own, although the difference manifests itself only after the antecedent time. Nonetheless, in the relevant sense the closest worlds where Nixon pressed the button are worlds where there is a nuclear holocaust – this is possible because the relevant notion of closeness is one which places little weight to differences of particular fact after the antecedent time. This notion of closeness is one that varies from antecedent to antecedent.

The upshot is that if we interpret $f(A, x)$ in the formal semantics as 'the A-closest A-world', where 'A-closesness' may depend on A, constraints F1-F3 remain intact while the problematic constraint F4 is no longer forced on us and CSO is not validated.

Stalnaker considers this response to Tichý's counterexample to CSO, writing: 'call a selection function which is based on some antecedent-independent ordering of possible worlds a *regular* selection function. The issue is whether one should describe the situation in terms of a single irregular selection function, or in terms of a contextual shifts from one regular selection function to another.' Stalnaker concedes, however, that the issue may be in part a matter of preference, not substance, regarding how best to distribute the burden of explanation between pragmatics and semantics. If this is so then, far from being the core of the Stalnakerian theory of conditionals, the principle CSO is more of a side issue. If CSO is a casualty of adopting a systematic epistemology of conditional statements then perhaps this is not so bad after all.

2.3.3 Conditional Logic without Stalnaker's constraint

What happens if we remove F4? Call the class of frames in which the selection function satisfies F1-F3 $\mathcal{C}$. Then $\mathcal{C}$ is sound and complete with respect to the logic CK with the addition of the principles ID, MP and CEM.[26] If we additionally stipulate that the frames satisfy

F5. If $f(A, x) \subseteq B$ and $f(B, x) = \emptyset$, $f(A, x) = \emptyset$.

then we also validate C1.[27] Note, however, that one can satisfy F5 without satisfying F4, and indeed one can have all the principles listed except for CSO in a semantics of the kind described.

The logic CK consists of the principles CK, RCN and RCEA. Call the result of adding ID, MP and CEM to the logic CK, L1 and let L2 be the logic resulting

[26] This is a simple application of the canonical model style of argument outlined in Chellas [6].

[27] C1 is a principle governing the conditions under which conditionals are vacuously true. F5 is roughly equivalent to saying that $f(A, x) = \emptyset$ just in case A is true in none of the worlds accessible to x (where y is accessible to x iff $f(\{y\}, x) \neq \emptyset$.)

from adding C1 to $\mathsf{L1}$. There is a sense in which $\mathsf{L1}$ is simply the logic you get when you drop CSO from Stalnaker's logic since $\mathsf{L1}+\ \mathsf{CSO}$ has exactly the same theorem's as Stalnaker's logic $\mathsf{C2}$.[28]

In order to compare $\mathsf{L2}$ to other systems of conditional logic note that, for example, $\mathsf{L2}$ is the result of replacing VLAS with CEM in Pollock's logic of conditionals SS. Removing CEM from $\mathsf{L2}$ and adding CSO, the principle $((\phi \to \psi) \wedge \neg(\phi \to \neg\chi)) \to (\phi \wedge \chi \to \psi)$ (a strengthening of VLAS) and $\phi \wedge \psi \to (\phi \to \psi)$ (which is already theorem of $\mathsf{L2}$) results in Lewis's VC (see [3].)

It is worth pointing out that adding any of CSO, CA, RCA, VLAS or LT to $\mathsf{L2}$ results in a system equivalent to Stalnaker's.[29] Thus out of logics consisting of the principles we have listed, $\mathsf{L2}$ is the strongest system containing CEM which does not collapse into Stalnaker's.

3 The tenability of CP

In this section we will construct a model for $\mathsf{CP}_\emptyset$ and CP within the logic $\mathsf{L2}$. While models of $\mathsf{CP}_\emptyset$ based on the semantics of section 2.3.1 seem worth investigating we shall limit our attention here to models based on the semantics we discussed in section 2.3.2, utilising an antecedent relative closeness ordering. We begin by defining the kind of model we are aiming for

Definition 3.0.1. *A probabilistic frame is a quintuple* $\langle W, F, P, f, \lambda\rangle$ *where*

- W *is a set (the set of worlds.)*
- $F \subset \mathcal{P}(W)$ *is a* σ*-algebra, i.e., a set containing the empty-set which is closed under complements in* W *and countable unions which is also closed under the operation* $X \Rightarrow Y := \{w \mid f(X, w) \in Y\}$.
- P *is a non-empty set of probability measures over* F. *I.e.* $Pr : F \to [0, 1]$ *for* $Pr \in P$.
- $f : \mathcal{P}(W) \times W \to W$ *is called the selection function.*
- $\lambda \notin W$ *is the impossible world.*

A probabilistic frame is adequate iff for every $Pr \in P$, $A, B \in F$, $Pr(A \Rightarrow B) = Pr(B \mid A)$ *whenever* $Pr(A) > 0$. *Here* $A \Rightarrow B := \{w \mid f(A, w) \in B\}$. *(It follows that in an adequate frame* $A \Rightarrow B$ *is measurable if* A *and* B *are, otherwise the left-hand side of this equation would not be defined.)*

In effect an adequate probabilistic frame consists of an interpretation of the conditional plus a set of ur-priors (the set P) which obey $\mathsf{CP}_\emptyset$ on measurable

[28] Lewis's axiomatisation of $\mathsf{C2}$ consists of $\mathsf{L1}$, CSO and the complicated principle $(A \vee B \to A) \vee (A \vee B \to B) \vee ((A \vee B \to C) \equiv (A \to C) \wedge (B \to C))$. The last principle is actually redundant (i.e. is provable in $\mathsf{L1}+\mathsf{CSO}$) but does not appear to be a theorem of $\mathsf{L1}$ (although it is a theorem of $\mathsf{L2}$.)

[29] For proofs of these facts see [XXXX], to appear.

sets. The more interesting frames are ones in which P represents a rich set of probability functions.

Intuitively there are two kinds of propositions: those which are not hypothetical at all, such as, for example, the proposition that a particular coin, C, landed heads, and those which are hypothetical, such as the proposition that if C is flipped it will land heads. The latter kind of proposition is often associated with a curious epistemic phenomenon: it does not seem to be possible to know whether the latter proposition is true if the coin isn't flipped. For example, if you accept conditional excluded middle then either C will land heads if it is flipped, or it will land tails, but in worlds where the coin is not flipped it is impossible to obtain further evidence to settle the question of which way it would land if flipped. Philosophers subscribing to the law of conditional excluded middle have conjectured that hypothetical propositions like this are a special source of indeterminacy (e.g. [33].) Whether or not this is so we can certainly agree that we must be ignorant in the scenario described, much as we would be in the face of vagueness or indeterminacy. In our model the non-hypothetical (completely determinate) propositions are represented by a Boolean algebra $\mathbb{B}$, which is a subalgebra of a larger space of propositions, $\mathbb{B}'$, consisting of all propositions hypothetical or not. The basic intuition is that one can have any credence you like regarding the completely determinate non-hypothetical facts, but once you have fixed your credences in those propositions your credences over the rest of the space of propositions is fixed. For example, if you know that C is fair and will not be flipped, then you are forced to have a credence of a half in the proposition that C will land heads if flipped. The situation here is similar to the analogous situation with vague propositions. Once you know someone has a certain borderline number of hairs, N, you are forced to be uncertain, to some degree, in the proposition that that person is bald.

3.1 The construction

Despite the impossibility results there have been a number of results to the effect that, under certain restrictions, one can have Stalnaker's thesis (see for example, van Fraassen [37], McGee [22], Stalnaker and Jeffrey [29], Kaufmann [15], Bradley [5].) However, in general, these constructions fall short of the full generality of a principle like $\mathsf{CP}_\emptyset$ as they do not deal with certain embedded conditionals. The closest thing to the project attempted here is found in van Fraassen's [37], where a model of $\mathsf{CP}_\emptyset$ is presented. In this paper van Fraassen provides, for each probability function, Pr, a model of the logic $\mathsf{L1}$ such that $Pr(A \to B) = Pr(B \mid A)$ whenever $Pr(A) > 0$. Unfortunately this construction is not suitable for our purposes. Van Frassen constructs a *different* conditional for each choice of probability function Pr, so this construction is not a satisfactory model of $\mathsf{CP}_\emptyset$ except in the limiting case in which there is only one 'Carnapian' ur-prior. Furthermore, it is not a model of the general principle CP. (There are also a number of technical limitations on the construction: (i) the construction only generates models where the number of measurable sets is countable (it is therefore also not a σ-algebra) (ii) the semantics is not based on selection functions and subsequently has an unnatural quantificational logic (for example the principle $\forall x(\phi \to \psi) \supset (\phi \to \forall x\psi)$, when x is not free in ϕ, is invalid)[30] (iii) the conditional operator is not defined on unmeasurable propositions (thus given (i), the conditional is only defined on countably many propositions) (iv) as far as I can see van Fraassen shows that $\mathsf{CP}_\emptyset$ holds over a field of sets, but not over the σ-algebra it generates.)

In that same paper van Fraassen provides a distinct model for a restriction of $\mathsf{CP}_\emptyset$ in which the antecedent and consequent, A and B, do not involve any conditionals themselves (see the 'Bernoulli-Stalnaker' models of [37].) However van Fraassen's construction does not allow iterated conditionals, and it validates the undesirable logic $\mathsf{C2}$. The approach here extends the idea of the 'Bernoulli-Stalnaker' models to validate the full strength of CP, iterations and all.

The construction begins with an initial set of possible worlds W, which represent maximally specific things that can be said about the world without mentioning conditional facts – facts about what will happen if this or that happens. The set W_∞ then extends this set, dividing members of W into epistemic possibilities according to the kind of hypothetical distinctions you can make. In the Bernoulli-Stalnaker models of van Fraassen we can think of W_∞ as pairs of members of W and well-orderings of the universe W. The well-ordering represents the closeness facts at that possible world. Even if we know exactly which world obtains you might be ignorant about the hypothetical facts, encoded by which well orderings obtain at that world.

Note that the space of worlds just described gives a poor representation of the conditional facts at a world since it only well-orders the possible worlds, not the epistemically possible worlds. What we want, rather, is a space of worlds W_∞ such that each member of W_∞ determines a well-ordering of W_∞, not W (actually, in our setting it's slightly more complicated because the ordering is antecedent dependent.) This is potentially problematic for cardinality reasons: W_∞ cannot be the same same size as the set of well-orderings over W_∞. To solve this we restrict our attention to well-orderings of some fixed collection of order types. For simplicity this is just ω.

Let us put this into practice. Assume that the initial set of states, W, that do not involve conditional facts is given and is countable. The set of worlds in our model will be the set $W_\infty = W^{\omega_1} = \{\pi \mid \pi : \omega_1 \to W\}$. Let $\mathbb{B}_\infty = \mathcal{P}(W_\infty)$.

Given our initial space W, define the following sequence of sets for $\alpha < \omega_1$

- $W_\alpha = W^{\omega^\alpha}$

That is, W_α represents the set of all ω^α sequences of members of W. Since $\omega^\alpha < \omega_1$ whenever $\alpha < \omega_1$ it follows that an element of W_α will be isomorphic to an initial section of a member of W_∞. Note also the following consequences of this definition:

- $W_0 = W$
- $W_{\alpha+1} \cong W_\alpha^\omega$

[30] Although, since this is effectively the limit assumption, perhaps this isn't so bad.

In what follows we shall adopt a practice of identifying products which are isomorphic to subsets of W_∞, allowing us, for example, to identify $A \times W_\infty$ with a subset of W_∞ whenever A is contained in some W_α.

The sets W_α for $\alpha < \omega_1$ help us describe the measurable sets.

Definition 3.0.2. *Suppose X is a set of subsets of W_∞. Then $cl(X)$ is the closure of X under the operations of countable unions and intersections, and complements relative to W_∞.*

The measurable sets, which we shall denote F_∞, can be thought of as being approximated by an infinite sequence of σ-algebras, $F_\alpha \subset \mathbb{B}_\infty$ for $\alpha < \omega_1$.

- $F_0 = \{A \times W_\infty \mid A \subseteq W\}$
- $F_{\alpha+1} = cl\{A_0 \times \ldots \times A_n \times W_\infty \mid A_i \times W_\infty \in F_\alpha \text{ for } 0 \leq i \leq n\}$
- $F_\gamma = cl(\bigcup_{\alpha<\gamma} F_\alpha)$

Note that $F_{\alpha+1}$ is generated by sets of the form $A_0 \times \ldots \times A_n \times W_\infty$ where $A_i \subseteq W_\alpha$. Each of these generating sets consists of an ω_1 sequence such that an initial finite number of elements belong to W_α and the rest belong to W. This is, of course, just equivalent to an ω_1 sequence of elements of W whenever $\alpha < \omega_1$: it is just equivalent to n successive ω^α-sequences of elements of W followed by an ω_1-sequence of elements of W, which is itself an ω_1-sequence of elements of W. Bearing this equivalence in mind we can see from the construction that an arbitrary member of F_α will be of the form $A \times W_\infty$ where $A \subseteq W_\alpha$. It is straightforward to show

Proposition 3.1. *$F_\alpha \subseteq F_\beta$ if $\alpha \leq \beta$*

Now we turn to our definition of F_∞, the set of measurable sets.

Definition 3.1.1. *A set $A \in \mathbb{B}_\infty$ is* measurable *iff $A \in F_\alpha$ for some α. We denote the set of measurable sets $F_\infty := \bigcup_{\alpha<\omega_1} F_\alpha$.*

It should now become apparent why we chose the ordinal ω_1 in our definitions: it is due to this choice that our measurable sets are closed under countable unions so that F_∞ is a σ-algebra.

Definition 3.1.2. *If A is measurable then the* rank *of A is the smallest α such that $A \in F_\alpha$. We shall write this: $rank(A) = \alpha$. If A is not measurable then $rank(A) = \infty$.*

For any ordinal $\alpha < \omega_1$ and ω_1 sequence, π, let $\pi[\alpha]$ be the ω_1 sequence starting from π's αth element. I.e. $\pi[\alpha](\beta) = \pi(\alpha+\beta)$. For each ordinal $\alpha < \omega_1$ and element $\pi \in W_\infty$ we can define a well-ordering of W_∞.

> $\tau \leq_{\alpha,\pi} \sigma$ if and only if for some $i, j \in \omega$ $\tau = \pi[\omega^\alpha.i]$, $\sigma = \pi[\omega^\alpha.j]$ and $i \leq j$.

We identify $\leq_{\infty,\pi}$ arbitrarily with $\leq_{0,\pi}$

It is now time to define the selection function for a $A \in \mathbb{B}_\infty$ of rank α (possibly identical to ∞)

$$f(A,\pi) = \begin{cases} \text{the `smallest' } A\text{-world under the ordering } \leq_{\alpha,\pi} & \text{if such a world exists} \\ \lambda & \text{otherwise} \end{cases}$$

By the 'smallest' A-world I just mean a world τ such that τ is in the domain of $\leq_{\alpha,\pi}$, $\tau \in A$ and for any other σ in the domain of $\leq_{\alpha,\pi}$ and in A, $\tau \leq_{\alpha,\pi} \sigma$. Note that, like in Stalnaker's semantics, $f(A,\pi)$ represents the closest world to π which belongs to A – the crucial difference is that the notion of closeness at play here depends on the antecedent, A – this accords with the semantics of section 2.3.2. Note that $f(A,\pi) \in A$ and $f(A,\pi) = \pi$ whenever $\pi \in A$. Thus this frame validates the logic L1.

Proposition 3.2. *$\emptyset$ is measurable and if A, B and $A_0, A_1, A_2, \ldots$ are measurable then so is $W_\infty \setminus A$, $A \Rightarrow B$ and $\bigcup_n A_n$.*

We now define the set, P, of ur-priors. For simplicity we shall assume that W is countable so that every subset of W can be treated as a measurable set (although it would be simple enough to drop this assumption and work with an initial σ-algebra over W instead.) We shall show that every regular countably additive probability function Pr on the powerset algebra on W extends to the measurable sets over $\mathbb{B}_\infty$. We then identify P with the set of all such probability functions generated this way.

Suppose that Pr is a regular countably additive probability function on $\mathbb{B}$. For $\alpha \leq \omega_1$ we define Pr_α over F_α as follows.

- $Pr_0 = Pr$
- $Pr_{\alpha+1}(A_0 \times \ldots \times A_n \times W_\infty) = Pr_\alpha(A_0 \times W_\infty) \ldots Pr(A_n \times W_\infty)$; $Pr_{\alpha+1}$ extends to the rest of $F_{\alpha+1}$ via Carathéodory's extension theorem.
- $Pr_\gamma(A) = Pr_\alpha(A)$ when $A \in F_\alpha$ for $\alpha < \gamma$. This extends to the rest of F_γ by Carathéodory's extension theorem.

Write Pr_∞ for Pr_{ω_1}. Observe, from the construction of Pr_∞, that for any $\alpha < \omega_1$ and $A_0, \ldots, A_k \subset W_\alpha$ $Pr_\infty(A_0 \times \ldots \times A_k \times W_\infty) = Pr_\infty(A_0 \times W_\infty) Pr_\infty(A_1 \times W_\infty) \ldots Pr_\infty(A_k \times W_\infty)$

We are now in a position to prove our main theorem.

Theorem 3.3. *The frame $\langle W_\infty, F_\infty, P, f, \lambda \rangle$ is adequate.*

In particular, if Pr is a countably additive regular probability function over W then $Pr_\infty \in P$ and $Pr_\infty(A \Rightarrow B) = Pr_\infty(B \mid A)$ whenever A and B are measurable.

Proof. Suppose that $rank(A) = \alpha$ so that $A = A' \times W_\infty$ for some $A' \subseteq W_\alpha$.

According to our definition $\pi \in A \Rightarrow B$ if and only if the smallest A world in the sequence $(\pi[\omega^\alpha.i])_i$ is a B world. In other words, if and only if $\pi[\omega^\alpha.0] =$

$\pi \in A \cap B$ or $\pi \notin A$ but $\pi[\omega^\alpha.1] \in A \cap B$ or $\pi[\omega^\alpha.0] \notin A$, $\pi[\omega^\alpha.1] \notin A$ and $\pi[\omega^\alpha.2] \in A \cap B$ or ... or $\pi[\omega^\alpha.i] \notin A$ for any i.

Let R be the set of π with $f(A, \pi) = \lambda$. Thus $A \Rightarrow B = (A \cap B) \cup (\overline{A'} \times (A \cap B)) \cup (\overline{A'} \times \overline{A'} \times (A \cap B)) \cup ... \cup R = \bigcup_n (\bar{A'}^n \times (A \cap B)) \cup R$. Here I am using $\overline{X}$ to denote the complement of X.

Note that $R \subseteq (\overline{A'})^\omega \times W_\infty$ which has probability 0 whenever $Pr_\infty(A) > 0$. Since we are calculating a union of disjoint sets we have

$$Pr_\infty(A \Rightarrow B) = \sum_{n<\omega} (Pr_\infty(\overline{A})^n \cdot Pr_\infty(A \cap B)) = \frac{Pr_\infty(A \cap B)}{1 - Pr_\infty(\overline{A})} = \frac{Pr_\infty(A \cap B)}{Pr_\infty(A)} = Pr_\infty(B \mid A)$$

□

It is fairly straightforward now to model CP. Let Γ be a set of propositions. Then we now define the selection function, f_Γ, for the connective $\rightarrow_\Gamma$ as follows.

Firstly we define $\leq_{\Gamma,\alpha,\pi} = \leq_{\alpha,\pi} \cap (\bigcap \Gamma \times \bigcap \Gamma)$. Suppose that $rank(\bigcap \Gamma \cap A) = \alpha$.

$$f_\Gamma(A, \pi) = \begin{cases} \text{the `smallest' } A\text{-world under the ordering } \leq_{\Gamma,\alpha,\pi} & \text{if such a world exists} \\ \lambda & \text{otherwise} \end{cases}$$

The next question we must settle is: what kind of proposition could be an agent's total evidence? A natural answer is suggested by our model: propositions that correspond to a set of possible worlds in W (i.e. a proposition of the form $A \times W_\infty$ for $A \subset W$.) A proposition that carves out finer distinctions than a set of possible worlds would is an essentially hypothetical proposition. For example, the proposition that a certain coin will land heads if flipped is a hypothetical proposition. It can certainly be *entailed* or *ruled out* by our evidence (for example, if we knew that the coin was flipped and landed heads or if we knew that the coin was flipped and landed tails.) In these cases, however, the proposition would not be our total (i.e. our strongest) evidence; it is simply impossible, I have argued, for our strongest evidence to be a hypothetical proposition.

If we let E be the conjunction of the evidence in Γ then our definition of $A \Rightarrow_\Gamma B$ above effectively amounts to: $A \cap E \Rightarrow B$. We shall now demonstrate that CP holds in this model.

Theorem 3.4. *If $E \in \mathcal{F}_0$ then $Pr_E(A \Rightarrow B) = Pr(AB \mid E) + Pr(\bar{A} \mid E)Pr(B \mid A)$ where $Pr_E(\cdot) = Pr(\cdot \mid E)$.*

Proof. Suppose that $E = E' \times W_\infty$ where $E' \subseteq W$. Then in general, for any α, and $A_0 \dots A_k \subseteq W_\alpha$, $(A_0 \times \dots \times A_k \times W_\infty) \cap E = (A_0 \cap (E' \times W_\alpha)) \times A_1 \times \dots \times A_k \times W_\infty$.

From theorem 3.3 we know that $Pr_E(A \Rightarrow B) = \Sigma_n Pr_E(\overline{(A')}^n \times AB) = \frac{1}{Pr(E)} \Sigma_n Pr((\bar{A})^n \times AB) \cap E)$.

Expanding this sum and applying the observation above we get $= \frac{1}{Pr(E)}(Pr(ABE) + Pr(\bar{A}E)\Sigma_n(Pr(\bar{A})^n Pr(AB)) = \frac{1}{Pr(E)} Pr(ABE) + \frac{1}{Pr(E)} Pr(\bar{A}E)Pr(B \mid A)) = Pr(AB \mid E) + Pr(\bar{A} \mid E)Pr(B \mid A)$ □

Now since $A \Rightarrow_\Gamma B = AE \Rightarrow B$ where E is the conjunction of Γ, we get $Pr(A \Rightarrow_\Gamma B) = Pr(ABE \mid E) + Pr(\overline{AE} \mid E)Pr(B \mid AE) = Pr(B \mid AE) = Pr(B \mid \{A\} \cup \Gamma)$.

Finally we must address the matter of the principle C1 which is not validated in this current construction, but is highly desirable. It is a simple matter, in fact, to modify the construction to validate C1. It is sufficient merely to ensure that the domain of $\leq_{\alpha,\pi}$, which we shall write $dom(\leq_{\alpha,\pi})$, is the same as the domain of $\leq_{\beta,\pi}$ for any α and β. For each world, π, let D_π be the union $\bigcup_\alpha dom(\leq_{\alpha,\pi})$. Then, for each α simply add all the elements of $D_\pi \setminus dom(\leq_{\alpha,\pi})$ on to the end of the ordering $\leq_{\alpha,\pi}$ in some order or other, it does not matter which, to make a longer ordering whose domain is D_π. The probability calculation goes through as in theorem 3.3 except in this case R denotes the set of worlds, π, where none of the first ω worlds (in the $\leq_{\alpha,\pi}$ ordering) is an antecedent world; as before this set has measure zero.

3.2 A small refinement of Hájek and Hall's theorem

Here we weaken the assumption CSO of [12] to VLAS. While VLAS and CSO are equivalent in L2, VLAS is strictly weaker in L1. Note that in the other direction CSO entails VLAS given CC, ID and CK.

To see that VLAS is strictly weaker in L1, here is a frame validating L1+VLAS but which does not validate L1+CSO. Let $W = \mathbb{N}$. If A contains primes then $f(A, i) = \{j\}$ where j is the smallest number $\geq i$ with $j \in A$, and $= \emptyset$ if there is no such number. If A does not contain i or any primes then $f(A, i) = \emptyset$, and if A contains i then $f(A, i) = \{i\}$. It's simple to verify that if $f(A, x) \subseteq B$ then $f(A \cap B, x) \subseteq f(A, x)$, ensuring the validity of VLAS. However, let $A = \{2\}$ and $B = \{2, 3\}$, so $f(A, 1) = \emptyset$ and $f(B, 1) = \{2\}$. So $f(A, 1) \neq f(B, 1)$ even though $f(A, 1) \subseteq B$ and $f(B, 1) \subseteq A$.

Theorem 3.5. *Suppose that $\mathsf{CP}_\emptyset$ holds for at least one ur-prior, Pr, and that the following facts are true of the model*

MP $A \cap (A \to B) \subseteq B$

CC $(A \to B) \cap (A \to C) \subseteq (A \to (B \cap C)$

VLAS $((A \to B) \cap (A \to C)) \subseteq ((A \cap B) \to C)$

Then there are no more than two disjoint consistent propositions.

Proof. Suppose AB, $A\bar{B}$ and $\bar{A}$ are three disjoint consistent propositions. They therefore all have positive probability by the regularity assumption. Let $C = A \cup (A \to B)$. Now note that the following chain of entailments hold (i.e. each proposition is a subset of the next):

1. $C \to AB$
2. $\subseteq (C \to AB) \cap (C \to A) \cap (C \to B)$ by CC

3. $\subseteq (C \to AB) \cap (CA \to B)$ by VLAS
4. $\subseteq (C \to AB) \cap (A \to B)$ since $CA = A$
5. $\subseteq (C \to AB) \cap C$ since $(A \to B) \subseteq C$
6. $\subseteq ABC$ by MP

So $C \to AB \subseteq ABC$ and thus $Pr(C \to AB) \leq Pr(ABC)$. Now the argument proceeds as in [12]:

1. $Pr(AB \mid C) \leq Pr(ABC)$, by $\mathsf{CP}_\emptyset$
2. So $Pr(C) = 1$ by probability theory.
3. $Pr(\bar{A} \cap (A \to \bar{B})) = 0$ by some fiddling (see P5 in [12].)
4. $Pr(A\bar{B}) = Pr(A \cap (A \to \bar{B}) = Pr(A \to \bar{B})$ by 3 and MP.
5. $Pr(\bar{B} \mid A) = Pr(\bar{A}B)$, i.e. $Pr(A\bar{B}) = Pr(A)(A\bar{B})$ which is impossible since $Pr(A) < 1$ and $Pr(A\bar{B}) > 0$.

□

In fact, both the Hájek and Hall result, Stalnaker's and the above can be seen as a special instance of an observation Stalnaker makes in [32].

Consider the following rule of proof:

RS If $\vdash (\phi \to \psi) \supset \phi$ then $\vdash \phi$.

If L is conditional logic that is, in a sense, not only sound but 'complete' for $\mathsf{CP}_\emptyset$ then SR must be an admissible rule in the sense of Lorenzen: i.e. the result of closing L under RS leaves you with L again, (or, in other words, RS is a derived rule of the logic.)

Theorem 3.6. *Assuming* MP*,* RS *is probabilistically valid in the sense that it preserves probability 1 in any model where* $Pr(A \to B) = Pr(B \mid A)$*, stipulating that* $Pr(B \mid A) = 1$ *when* $Pr(A) = 0$*.*

Proof. Suppose that $Pr((A \to B) \supset A) = 1$. Then $Pr(A \to B) \leq Pr(A \cap (A \to B))$ by probability theory, which is $\leq Pr(A \cap B)$ by MP.

Since $Pr(A \to B) = Pr(B \mid A)$, $Pr(B \mid A) \leq Pr(A \cap B)$ so that $Pr(A) = 1$ □

However, RS is not admissible in C2, nor MP+CC+CSO, nor MP+CC+VLAS. In the former case the system you get by closing under RS collapses $\to$ into the material conditional (which is incompatible with $\mathsf{CP}_\emptyset$ and the existence of three disjoint consistent propositions), and similarly for the latter two systems if we add CEM or $(\phi \wedge \psi) \supset (\phi \to \psi)$ to either (which are both guaranteed by $\mathsf{CP}_\emptyset$ anyway.)[31]

[31] Since we can prove $(C \to A \wedge B) \supset C$, where $C = A \vee (A \to B)$, it follows that $A \vee (A \to B)$ is a theorem for arbitrary A and B. Since $A \supset B, \neg A \vdash A \to B$, and $A \supset B, A \vdash A \to B$ we have $A \supset B \vdash A \to B$ (the converse is MP.)

References

[1] E.W. Adams. *The logic of conditionals: An application of probability to deductive logic*, volume 86. Springer, 1975.

[2] A. Ahmed. Walters on conjunction conditionalization. In *Proceedings of the Aristotelian Society (Hardback)*, volume 111, pages 115–122. Wiley Online Library, 2011.

[3] H. Arlo-Costa. The logic of conditionals. 2007.

[4] J.F. Bennett. *A philosophical guide to conditionals*. Oxford University Press, 2003.

[5] R. Bradley. Multidimensional possible-world semantics for conditionals. *Philosophical Review*, 121(4), 2011.

[6] B.F. Chellas. Basic conditional logic. *Journal of philosophical logic*, 4(2):133–153, 1975.

[7] C.B. Cross. Antecedent-relative comparative world similarity. *Journal of Philosophical Logic*, 37(2):101–120, 2008.

[8] V.H. Dudman. Conditional interpretations of if-sentences. *Australian Journal of Linguistics*, 4(2):143–204, 1984.

[9] D. Edgington. Counterfactuals and the benefit of hindsight. *Causation and Counterfactuals. London: Routledge*, pages 12–27, 2004.

[10] A. Gibbard. Two recent theories of conditionals. *Ifs: Conditionals, Belief, Decision, Chance and Time, WL Harper, R. Stalnaker & G Pearce (eds.). D. Reidel Publ. Co., Dordrecht*, pages 211–248, 1981.

[11] A. Hájek. *Most counterfactuals are false.* MS.

[12] A. Hájek and N. Hall. The hypothesis of the conditional construitl of conditional probability. *Probability and conditionals: belief revision and rational decision*, page 75, 1994.

[13] J. Hawthorne. Chance and counterfactuals1. *Philosophy and Phenomenological Research*, 70(2):396–405, 2005.

[14] L. Humberstone. *The connectives*. MIT Press, 2011.

[15] S. Kaufmann. Conditional predictions. *Linguistics and Philosophy*, 28(2):181–231, 2005.

[16] I. Kvart. *A theory of counterfactuals*. Hackett Publishing Company, 1986.

[17] D. Lewis. Counterfactuals oxford, 1973.

[18] D. Lewis. Probabilities of conditionals and conditional probabilities. *The philosophical review*, 85(3):297–315, 1976.

[19] D. Lewis. Counterfactual dependence and time's arrow. *Noûs*, pages 455–476, 1979.

[20] J. Mårtensson. Subjunctive conditionals and time: A defense of the classical approach, 1999.

[21] V. McGee. A counterexample to modus ponens. *The Journal of Philosophy*, 82(9):462–471, 1985.

[22] V. McGee. Conditional probabilities and compounds of conditionals. *The Philosophical Review*, 98(4):485–541, 1989.

[23] C.J.G. Meacham. Two mistakes regarding the principal principle. *The British Journal for the Philosophy of Science*, 61(2):407–431, 2010.

[24] D. Nolan. Defending a possible-worlds account of indicative conditionals. *Philosophical Studies*, 116(3):215–269, 2003.

[25] P. Noordhof. Tooley on backward causation. *Analysis*, 63(278):157–162, 2003.

[26] J.L. Pollock. *Subjunctive reasoning*. Reidel Dordrecht, 1976.

[27] J.L. Pollock. A refined theory of counterfactuals. *Journal of Philosophical Logic*, 10(2):239–266, 1981.

[28] J. Schaffer. Counterfactuals, causal independence and conceptual circularity. *Analysis*, 64(284):299–308, 2004.

[29] Jeffrey R. Stalnaker, R. Conditionals as random variables. *Probability and conditionals: belief revision and rational decision*, page 31, 1994.

[30] R. Stalnaker. A theory of conditionals. *Studies in logical theory*, 2:98–112, 1968.

[31] R. Stalnaker. Indicative conditionals. *Philosophia*, 5(3):269–286, 1975.

[32] R. Stalnaker. Letter to van fraassen. *WL Harper and CA Hooker (1976)*, pages 302–306, 1976.

[33] R. Stalnaker. A defense of conditional excluded middle. *Ifs: Conditionals, belief, decision, chance, and time*, pages 87–104, 1981.

[34] R.C. Stalnaker. Probability and conditionals. *Philosophy of Science*, pages 64–80, 1970.

[35] P. Tichỳ. A new theory of subjunctive conditionals. *Synthese*, 37(3):433–457, 1978.

[36] M. Tooley. Backward causation and the stalnaker-lewis approach to counterfactuals. *Analysis*, 62(275):191–197, 2002.

[37] B.C. Van Fraassen. Probabilities of conditionals. *Foundations of probability theory, statistical inference, and statistical theories of science*, 1:261–308, 1976.

[38] T. Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.