On the Semantics of Indicatives*

Abstract

In this paper I criticise orthodox possible worlds analyses of indicative conditionals that invoke the idea of a minimal deviation from actuality. Drawing on a recent discussion by Moritz Schulz I explore an alternative possible worlds semantics for indicatives based on the theory of random selection and argue that it is both superior and makes possible a genuinely probabilistic theory of conditionals.

In the mid to late 60s a couple of discoveries were made that significantly advanced our understanding of conditionals. One of these was the observation that it's possible to model the probabilities of conditional statements quite closely using the mathematical theory of conditional probability, where the probabilities in question represent rational degrees of belief or epistemic probabilities (Adams [1]). The other was the discovery of an abstract possible world semantics for conditionals based on the notion of a selection function, giving us an insightful look at the structure of the logic of conditionals (see Stalnaker [34] and Chellas [6]). Indeed, the logic of conditionals one obtains, when analysed in terms of the theory of selection functions and the probability calculus alike, coincided perfectly.[1] This suggested that these two ideas were not completely disconnected.[2]

These observations gave rise to a fairly shortlived research program – epitomized in Stalnaker's 1970 paper 'Probability and Conditionals' – that attempted

*THANKS

[1]See Gibbard [14] and Stalnaker [38]. There are some subtleties to do with these results that we shall get to later. For example, Gibbard is interested in Adams' probabilistic notion of validity, that concerns the preservation of 'high probability' for the portion of the language to which Adams' theory is applicable. There can in principle be rules preserve high probability that when reformulated as axioms might not have high probability (for example the rule $A \vdash B$ might be probabistically valid, whilst $\vdash A \supset B$ probabilistically invalid. However, since A or B might involve conditionals, $A \supset B$ would be outside the purview of Gibbards' theorem, which only treats unembedded conditionals).

[2]Of particular note was the fact that both the theory of selection functions and the probability calculus provided concrete models of the puzzling phenemona of non-monotonicity. Monotonic theories of conditionals have the rule $A \rightarrow C \vdash AB \rightarrow C$. The fact that the sentence 'if Alice went to the beach she had a good time' does not seem to entail 'if Alice went to the beach and got attacked by a shark she had a good time' suggests that natural language indicatives are non-monotonic. Prior to the two approaches mentioned above, many people had attempted to understand conditionals by analogy to kinds of inferences, and consequently had great difficulties explaining this phenoma: if the inference from A to B is good, adding more premises won't change this.

to unify the probabilistic and possible worlds theories.[3] However the program ground to a halt when a whole slew of mathematical results were proved that apparently showed that no truth-conditional theory (such the possible worlds theory) could deliver the probabilities predicted by the probabilistic theory (see Hajek and Hall [18] for a good overview of these results). Elsewhere I have argued that these results have been overblown somewhat, and that a particularly natural connection between the probabilistic and truth conditional theory survives (see section 1.3 of this paper and [REF]).[4]

In my view Stalnaker's original program was flawed in its attempt to incorporate two fundamentally different and competing ideas. The first was the idea that an indicative's truth conditions can be specified in terms of the notion of a minimal deviation from actuality: it is true if its consequent is true in a world where the antecedent is true and which deviates from actuality as little as possible. This idea is naturally modeled by a selection function semantics in which the selection function is interpreted as the most similar A-world to the world of evaluation.

The other idea is based on the probabilistic motivations outlined above, in which the probability that a conditional is true is given by the probability that the consequent is true given that the antecedent is true. It turns out that this idea can also be given a precise formulation in terms of a selection function, although a formulation that is in direct tension with the similarity theoretic interpretation of the selection function. On the alternative interpretation the selection function doesn't select the most similar accessible A-world, it selects an accessible A-world at random, and moreover selects in such a way that the probability of a world being selected is always proportional to the probability of that world (an idea which we will make precise later).[5]

In later work Stalnaker dropped the probabilistic theme and developed the similarity idea into a fully fleshed theory of indicatives. However, the other interpretation is still viable, and provides a formally similar possible worlds theory in which the constraints on the selection function are slightly different (and which therefore validates a slightly different conditional logic). In this paper I will argue that there are actually several problems that a similarity interpretation of indicatives faces – not all of which are directly related to probabilities – that the probabilistic interpretation is uniquely placed to address.

[3] See Stalnaker [38], van Frassen [44], Skyrms [32], Harper [19]. Although enthusiasm for the program has waned in the last two decades, there have been some more recent proponents as well such as, for example, Bradley [4].

[4] There are other philosophers who have questioned these results, and have proposed some weakened form of the connection between conditionals and probabilities (for a recent attempt see, for example, Bradley [4]). Most of these authors concede some casualties to the triviality results, usually by telling a special story about the probabilities of iterated conditionals; the theory I prefer in [REF] makes no such concessions.

[5] A similar, although importantly different interpretation is suggested in Lewis [27] (see analysis $1\frac{1}{2}$ in §1): a subjunctive conditional is true if the consequent is true at a certain antecedent world that is selected arbitrarily from among the closest accessible antecedent worlds. This analysis has recently been defended, also in the context of subjunctive conditionals, by Moritz Schulz [31]. By contrast, my analysis of indicatives does away with the notion of 'closeness' altogether, and unlike Schulz's theory, fully supports a version of Stalnaker's thesis.

I begin, in section 1, by outlining the selection function semantics in more detail, and show how the idea of random selection can be made precise and explain exactly how it delivers the correct connection to conditional probabilities. In section 2 I turn to the similarity interpretation. In section 2.1-2.3 I argue that for certain kinds of indicative conditionals the principle of conditional excluded middle, and the principle that conjoining things which are each true if A, gives you something that is true if A are both valid, and that their joint validity poses a special problem for an interpretation of indicatives in terms of similarity that is addressed in the probabilistic interpretation. In section 2.4-2.6 I look at a principle of conditional logic that is distinctive to the similarity interpretation and argue that it has some surprising consequences.

Before we move one, let me spend a moment delineating the scope of this paper. First of all, it should be noted that my arguments will primarily be directed at similarity accounts of *indicatives*: I will for the most part leave it open whether subjunctives can be given an adequate treatment in terms of similarity. Even among those who happily apply the similarity theory to the subjunctive case, the idea that this analysis extends to indicatives is, I think, already under some suspicion (perhaps the most prominent example of this attitude is David Lewis, who despite pioneering the similarity analysis of subjunctives, maintained that indicatives required a fundamentally different treatment).

Secondly, I will focus my attention on bare indicatives in the simple past tense, such as the sentence 'if Oswald didn't kill Kennedy someone else did'. I am thus excluding indicatives with modals in the consequent from consideration, such as the sentences 'if Oswald didn't kill Kennedy, someone else will/must/ought to/should/would have/...'. The interaction of conditionals with modals of various sorts is a complicated business, but an issue that is perhaps best resolved once we have first understood bare indicatives.[6] I shall also be setting aside conditionals in the simple present, such as the conditional 'if this coin is flipped it lands heads' – sentences in the simple present typically involve some covert habitual operator, and thus fall naturally into the second category.

1 Possible world semantics for indicative conditionals

Although indicatives present a particularly hard case for a truth conditional treatment, possible worlds accounts of indicatives have been defended by a few

[6]It should be noted, though, that some people, following Kratzer [23], take the opposite attitude: that we should start with the connection between modals and conditionals, and treat the case of bare indicatives as a special case. Such views do not treat conditionals as connectives, but as devices for restricting modal operators; the notion of a genuinely bare indicative doesn't really make sense in this theory – at the level of logical form apparently 'bare' indicatives end up having an invisible operator in the consequent. However the kinds of reasons motivating Kratzer to take the interaction between conditionals and modals as the starting point can be explained in ways that take conditionals to be connectives (see Dorr and Hawthorne [7]).

philosophers (see Stalnaker [35], Nolan [30] for example), and are quite pervasive in linguistics (a tradition inspired especially by Kratzer [23]).

The scope of this paper is of necessity relatively narrow. There are many general challenges to a possible worlds account of indicatives that I shall not be addressing. I take it that any adequate theory must provide explanations of the seeming validity of 'or-to-if' inferences (see Bennett [3] §18 for an overview), the 'import-export' principles, and an account of Gibbard cases (Gibbard [14]). Although I do not intend to defend the position here, it is my view that these puzzles can all be explained by paying careful attention to the context sensitivity of indicative conditionals. Theorists, like myself, who take this line usually invoke a contextualist framework in which the proposition expressed by an indicative conditional depends on the evidence of some salient individual or group. To constrain the subsequent discussion, I shall limit myself here to outlining, in very general terms, a particular contextualist framework in which these kinds of challenges might be met.[7] With a concrete framework on the table, we can focus our attention to the interpretational issues.

Abstracting from the interpretational issues for the moment, the contextualist framework I take to be the most promising has the following three key elements:

1. A set of *indices*, W.

2. An accessibility relation over W, E, called the *modal base*.

3. A function f, called the *ur-selection function*, that takes as arguments a set of indices A and an index x and outputs another index $f(A, x)$.

The indices can be interpreted in various ways depending on your preferred framework, but it is usual to think of them either as possible worlds or epistemic possibilities of some sort. Sets of indices are identified with propositions.

The accessibility relation, E, represents the evidence of someone or some group of people that is relevant to the conversation. For convenience we write $E(w)$ to denote the set of indices accessible to w. The modal base is sometimes equivalently given by a function mapping each index to a set of indices – the function mapping w to $E(w)$. The proposition $E(w)$ represents the total evidence possessed by the salient individual or group at w.

The pivotal element of this framework – the part that treats conditionals specifically – is the ur-selection function, f. If A is a set of indices and x an index, we write $f(A, x)$ (or sometimes $f_A(x)$) for the index that the selection function f outputs on these arguments. We typically require that f satisfy the following constraints:

ID $f(A, x) \in A$

MP $f(A, x) = x$ if $x \in A$

[7] Kratzer [23], van Rooij [45], Nolan [30].

The former corresponds to the law of identity, $A \to A$, and the latter to modus ponens for $\to$: $(A \to B) \supset (A \supset B)$. When I talk of a selection function in what follows I shall assume that it satisfies these two conditions.[8]

The interpretion of the selection function will depend on your application. For subjunctives we shall understand it neutrally as follows:

> $f(A, x)$ is the world that would have obtained if it had been the case that A (as evaluated at x).

Since our primary concern are indicatives it is more natural to think of the indices as describing the different possible ways things are for all you know. One of these describes the way things actually are: x. The indicative selection function is thus understood as follows:

> $f(A, x)$ denotes the way things are if A (as evaluated at x).

Some authors allow a slightly more general notion of selection function, in which the selection function outputs a set of indices instead of a single index: I shall call these 'Chellas selection functions' (see Chellas [6]). On this interpretation you can think of the set of worlds that the subjunctive ur-selection function outputs, for example as the worlds that might have obtained had A obtained.[9] This small tweak to the theory generalises the theory considerably, allowing one to model logics that do not satisfy the law of conditional excluded middle (discussed in section 2.2).

With this notion in hand we are in a position to give a semantics for a conditional statement of the form $A \to B$:

(*) $A \to B$ is true at world x iff B is true at $f(A, x)$

Informally, the conditional 'if A then B' is true iff B is true at the world that obtained if A. It is extremely natural, in this context, to engage in a bit of notational overloading. In what follows I shall use f_A (without an argument) as a singular term denoting the world that obtained if A. f_A is a (non-rigid) singular term whose denotation at a world x is just $f_A(x)$.

The ur-selection function, and the conditional it defines, does not straightforwardly correspond to ordinary natural language conditionals. The *ur*-selection function corresponds to the kind of conditional expressed in a context in which there is no evidence, or in which the salient evidence is completely vacuous. Indeed, the ur-selection function stands to ordinary everyday conditionals in a way that's somewhat analogous to the way that rational 'ur-priors' stand to the rational credences of informed people: to get an ordinary conditional, as expressed by someone whose evidence is represented by E, we have to 'conditionalise' the ur-selection function on E to get another selection function f_E

[8] I have glossed over the case where A is empty – in this case we can have f_A output the 'impossible world' at which every sentence is stipulated to be true (see Stalnaker [34]).

[9] This gloss seems reasonable under the assumption that 'might' counterfactuals dualize 'would' counterfactuals. A different gloss should be sought without that assumption (e.g. each member $f(A, x)$ is a world such that it's not the case that it wouldn't have obtained had A obtained).

which we then use to evaluate a conditional in which E is the salient evidence. The process for conditionalising a selection function is given by the following equation:[10]

Harper Updating: $f_E(A, x) = f(A \cap E(x), x)$

In other words, the truth conditions for an utterance of 'if A then B', when the salient evidence is represented by the accessibility relation E, is that the *accessible* A-world that gets ur-selected is a B-world. It is easy to check that if E is reflexive, f_E also satisfies the constraints ID and MP. Let us write $A \to B$ to denote the ur-conditional – the connective one gets from the ur-selection function – and write $A \to_E B$ for the conditional that corresponds to the updated selection function $f_E(\cdot, \cdot)$. According to the present contextualist picture an utterance of a conditional, 'if A then B' in a context where E is salient expresses the proposition $A \to_E B$.[11]

This is the abstract framework we will be assuming in what follows. One thing to emphasize about this formalism is that evidence, in this framework, is not represented by a salient proposition but rather by a function from worlds to propositions. According to a simplistic albeit more common variant of contextualism, the modal base is replaced by the proposition that constitutes the speakers evidence, or perhaps some other salient evidence, in the actual world (see for example the 'context set' of Stalnaker [40], chapter 4). The prevailing criticism of contextualism – which is really a criticism of the simplistic version of the theory – is that it is untenable because in order for Bob to know what Alice has said using a conditional sentence, Bob has to know what Alice's total evidence is, and this rarely, if ever, happens (Gibbard [14], Bennett [3], Stalnaker [39]). The present version of the theory is not subject to this criticism: if Bob knows that Alice is speaking, he knows the relevant modal base is one

[10] The rule was proposed first in Harper [19].

[11] Two contentious inferences are invalidated by the present semantics: the 'or-to-if' inference, $\neg A \vee B \vdash A \to B$, and the import-export law, $AB \to C \dashv\vdash A \to (B \to C)$. The fact that these principles cannot be derived in the selection function setting is both good and bad news. It's good news since both principles collapse the conditional into the material conditional given a small amount of background logic. The bad news is that both or-to-if and import-export seem valid for natural language indicatives, and so the appearance of validity needs explaining.

In fact, the ordinary selection function semantics invalidates a more general principle that allows us to derive both of the aboveprinciples – the rule of conditional proof with side premises: If $\Gamma, A \vdash B$ then $\Gamma \vdash A \to B$. For example, 'or-to-if' follows from conditional proof and the propositional law $\neg A \vee B, A \vdash B$. Similarly the fact that $AB \to C, A, B \vdash C$ gets us one direction of import-export by two applications of conditional proof, and the fact that $A \to (B \to C), AB \vdash C$ gets us the other direction.

In the contextualist framework, however, we have weakened variant of conditional proof: If $\Gamma, A \vdash B$ then $\Gamma \vdash A \to_\Gamma B$ where $\to_\Gamma$ denotes the result of updating the salient evidence by the propositions in Γ (E updated by a proposition A is $(E \cap A \times A) \cup (E \cap \overline{A} \times \overline{A})$). The contextualist version of conditional proof allows one to prove subscripted variants of these principles (for example, we can prove $\neg A \vee B \vdash A \to_{\neg A \vee B} B$). In my view these variant principles can explain the seeming validity of or-to-if and import-export, appealing to the fact that the premises of these arguments are usually salient evidence. Defending this claim in full would take us too far afield however.

that maps the world x to Alice's evidence at x. Bob does not need to know which world actually obtains, or what Alice's evidence is in the *actual* world, he merely needs to know who is speaking in order to work out what has been said.[12]

1.1 Interpreting the selection function

At this juncture one might object that the analysis of conditionals in terms of selection functions is not a particularly informative one. The selection function itself was specified in conditional language – after all, we defined f_A as 'the world that obtained *if* A'. Acknowledging this point does not diminish the value of the abstract semantics as a model theory – a mathematical tool allowing us to settle questions of validity, consistency and the like. However it does mean that we can't directly retrieve an informative account of conditionals from the formalism.

One might think that the theory of selection functions is substantive in its effect on the kind of logic it imposes on us. However, it turns out that even the assumption that one can represent natural language conditionals by a selection function is surprisingly weak. If a conditional connective, $\to$, satisfies a weak and relatively uncontroversial minimal logic called CK (analogous to the minimal modal logic K)[13], then there is a set of indices and Chellas style selection function on those indices that represents that connective: for any A and B, $A \to B$ is true if and only if it is true at the designated index of the selection function model.[14] For example, the material conditional satisfies this logic: the selection function representing it treats every false proposition as though it were inconsistent (i.e. outputs the empty set whenenever A is false), and outputs the singleton of the actual world when A is true. A strict conditional is represented by the selection function that outputs the accessible A-worlds, and so on. It follows, given these minimal assumptions, that there is this thing – the indicative selection function – out there. What is in dispute is not its existence, but the correct way to interpret it.

[12] The version of contextualism that uses a modal base rather than a simple proposition, derives from Kratzer's general theory of modals [23] (which in turn extends the logicians treatment of modality using accessibility relations, see Kripke [24]). The idea that this formalism also helps with this objection to contextualism is spelt out more fully in [XXX] [REF].

[13] The minimal logic has three principles. RCEA ensures that conditionals are not hyperintensional in the antecedent position: is says that if A and B are logically equivalent, $A \to C$ and $B \to C$ are logically equivalent. The principle RCN states that if B is a theorem so is $A \to B$. Finally the principle CK states $(A \to (B \supset C)) \supset ((A \to B) \supset (A \to C))$. In fact the analogy with the minimal modal logic K is closer than it might at first appear: the last two principles RCN and CK effectively tell us that the unary operator $A \to$ satisfies the two principles governing the minimal modal logic K.

[14] This result applies to any simple propositional logic expanded with a binary connective $\to$, and is a consequence of the completeness of CK with respect to a class of selection functions. Once one expands the language to contain infinitary conjunctions or quantifiers one has to start worrying about the limit assumption. If one assumes the principle LIM a similar representation theorem is possible, without it one has to generalise ones notion of a selection function (see footnote 15).

One might quite reasonably hope for a more illuminating way to characterise the selection function. Indeed this is exactly what Stalnaker and Lewis provide with their analysis of counterfactual conditionals in terms of the notion of *similarity*.[15] According to this idea, the selected A-world, $f(A, x)$, is to be thought of as the A-world, or set of A-worlds, *most similar* to actuality. Combining this with (*) delivers the following truth conditions for counterfactuals.

> SIMILARITY: If it had been the case that A it would have been the case that B iff f_A is a B-world, where f_A is the most similar (accessible) A-world to the actual world.[16]

If there is no such world the conditional is simply stipulated to be true. I have assumed above that there is a unique closest A-world; if there is more than one, we may modify the clause so that a subjunctive is true iff the consequent is true at all the most similar (accessible) A-worlds.

Analyses of conditionals based on similarity in one form of the other have dominated in the philosophy of conditionals in the last forty years or so. Recently, however, Moritz Schulz has defended an alternative way to understand the selection function semantics based on the notion of random selection (see Schulz [31]; the view is also considered, but not endorsed, by Lewis [27] §1). In what follows I want to suggest an interpretation, inspired by Schulz's view,[17] in which we think of the selected A-world as having been selected essentially at random from the (accessible) A-worlds. The selection might favour worlds that are more probable, but does not give overiding preference to worlds that are more similar to the actual world. The truth conditions (now applied to indicatives) becomes:

> RANDOM: If A then B iff f_A is a B-world, where f_A is a randomly selected (accessible) A-world.

This proposal initially seems to raise more questions than it answers. What does it mean for a world to be randomly selected? Is the notion of randomness based on objective chance or is it a more epistemic notion? For Schulz, who is primarily interested in providing a semantics for subjunctive conditionals, the notion of randomness is associated with objective chances. Schulz is happy to take the notion of random selection as bedrock, appealing to a metaphysically substantive theory of arbitrariness to ground these statements.[18]

[15] Actually, there is a technical reason relating to the limit assumption that means that Lewis's semantics cannot strictly be represented by a selection function, even of the general sort developed by Chellas. It is possible to introduce a very general selection function semantics that deals with this complication, but little turns on it so I shall supress the matter in what follows. (In the generalised semantics, the selection function outputs a filter over the set of indices instead of an index or set of indices; I develop this further in [REF].)

[16] Stalnaker [34], Lewis [25].

[17] Although note that my proposal is not strictly analogous to Schulz's account of counterfactuals for reasons mentioned is footnote 5.

[18] I am supressing many details of Schulz's account here. Schulz puts special emphasis on the epsilon calculus as a way of formalising the notion of arbitrary reference, for example, and Schulz evaluates a counterfactual by randomly selecting a world from among the set of closest

Although I am sympathetic to Schulz's analysis of subjunctives, I shall mostly set that matter to one side. In this paper I wish to suggest instead that the above interpretation is particularly well suited to providing a much-sought-after possible worlds style account of indicative conditionals. However, before that can succeed more needs to be done in explaining exactly what it means to say that a world has been selected 'at random', and to explain how we can talk about the probability of some world being selected. Once this is done, I shall argue that the result is exactly the kind of theory sought by people like Robert Stalnaker in 'Probability and Conditionals' – a theory which unifies the possible worlds selection function semantics with the probabilistic theory of conditionals.

1.2 Random selection

Let's begin with an informal gloss. According to the random world interpretation of the ur-selection function, f_A is to be interpreted as a world that has been selected from among the A-worlds at random. Just as some worlds are more similar to actuality than others, some are more probable than others: we select with a preference for more probable worlds, just as the similarity account has a preference for similar worlds. However unlike the similarity account, there is no *overriding* preference for the most probable A-world, rather, the more probable a world is, the more likely f_A is to pick it. The basic tenets of the random world interpretation can thus be summarized as follows:

1. f_A denotes a world that has been selected at *random* from the A worlds (usually no world has a probability 1 of being selected).

2. The probability that a given A-world has been selected is directly proportional to that worlds probability. More generally the probability that the selected world is in a subset B of A is proportional to the probability of B.[19]

These tenets are supposed to be understood as applying to our ur-selection function. Once the interpretation of the ur-selection function is fixed, this fixes the interpretation for each updated selection function: one selects a random *accessible* A-world instead.

This informal gloss of the interpretation calls out for clarification in several places. (i) What does it even mean for a function to pick a world at random, and how exactly does it get 'picked'? (ii) What kind of probability are we talking

(or most relevant) A-worlds. These differences are worth discussing further, but I think the basic idea of interpreting conditionals in terms of random selection is close enough to Schulz's to warrant the comparison.

[19] Note that the second thesis marks an important departure from a similar proposal due to Lewis and Schulz mentioned earlier. In the Lewis-Schulz version of the theory, f_A denotes a world that has been selected at random from among the *closest* A worlds, and so A worlds that are not among the closest A worlds have no chance of being selected even if they are quite probable. This feature makes the theory unsuited to capturing intuitive judgments about the probability of conditionals.

about? (iii) How do we make sense of the probability that a particular world has been selected?

To formulate and address these questions in a precise way, it will be helpful to introduce the concept of an *A-valued random variable*.[20] To illustrate, suppose that A is the set of balls contained in a particular bag, and we know that through some random process one of the balls in A is going to be selected: lets say I'm just going to stick my hand in the bag without looking and pick one. In this context we could introduce a term, b_A, meaning roughly 'the ball that's going to be selected by this process'. The term 'b_A' here is an example of a random variable. With this term introduced, we can then go on to ask, for any ball a in A whether $b_A = a$, or for any subset B of A, whether b_A belongs to B, and we can also ask for the probabilities of these hypotheses.

Formally an A-valued random variable is just a (potentially non-rigid) designator whose potential designata lie in the value space A. In what follows I shall write x_A, with the subscript indicating the value space of the random variable. The value space, A, can be any set, but for technical reasons we assume it always comes equipped with a sigma-algebra, Σ, of *measurable* subsets of A. Putting this together, x_A is a random variable iff:

> x_A is a function from possible worlds to A; i.e. a non-rigid designator whose candidate referents lie in the set A.
>
> x_A is a measurable function: $\{w \mid x_A(w) \in B\}$ (written $x_A^{-1}(B)$) is a measurable set of worlds whenever B is a measurable subset of A.

For the purposes of exposition, the second condition can mostly be set aside (it is relevant only when some of the sets involved are infinite and pathological). In our example where the value space A is the finite set consisting of balls in the bag, every subset of A is measurable, and b_A is a function that maps each world to the ball that is picked out of the bag at that world.

Given a probability function defined over measurable sets of worlds, Pr, and a random variable, x_A, we are in a position make sense of the probability of statements like $x_A = a$ and $x_A \in B$, where a is a member of A, and B a measurable subset of A. They are simply given by the probability of the set of worlds where the value of x_A is a or belongs to B respectively.

$$Pr(x_A = a) := Pr(\{w \mid x_A(w) = a\})$$

$$Pr(x_A \in B) := Pr(\{w \mid x_A(w) \in B\})$$

[20] Technically speaking, an '(A, Σ)-valued random variable' is a measurable function from the set of worlds to an arbitrary set A, with an associated sigma-algebra Σ defined over it. Some mathematicians, however, reserve the word 'random variable' (without qualification) for something that has a real number as a value with its usual sigma algebra, and use the word 'random element' for the more general concept. The approach of Jeffrey and Stalnaker [33] is an example of a view that treats conditionals in terms of the more narrow notion of random variable. It is important to note that this view therefore bears very little resemblence to the approach outlined here, despite also employing the notion of a random variable: the values of these variables are real numbers and not possible worlds (and the theory is not fully compositional as a result).

(Note that the corresponding sets of worlds are measurable and have probabilities provided $\{a\}$ and B are measurable subsets of A.)

In the application we are interested in the values of the random variables are not balls, but possible worlds themselves. If A is a set of worlds, then it is automatically equipped with a collection of measurable subsets inherited from the background space of worlds. A random variable, x_A, whose values lie in a set of worlds A, is by definition a function that maps each possible world to a possible world belonging to A, $x_A(x)$. This, of course, corresponds exactly to our definition of a selection function. Thus with this formalism in place, we are in a position to introduce a special set of random variables: for each set of worlds, A, we introduce the random variable f_A to denote the world that obtained if A obtained. In our framework it is perfectly sensible to ask things like 'what is the probability that the selected world, f_A, is the world w' or 'what is the probability that the selected world, f_A, belongs to B' and so on. The upshot, of course, is that we can give precise meaning to our earlier claims about the probability of this or that world being selected.

By convention, a random variable is called a random vector if its values are vectors, a random field if its values are fields, and so on; in keeping with this convention f_A is a random A-world, and a selection function is just a collection of random worlds – a random A-world for each set A.

Now let us return to the interpretive difficulties we encountered with the idea of a random selection. The idea of randomly selecting a ball from a bag is a useful analogy, but invites some misconceptions that need to be dispelled. To begin with, the formalism of random variables does not require the selection process to be a physical one such as the process of picking a ball out of a bag; it can be applied to any variable whose value is unknown. To illustrate, consider the following different procedures for selecting an element from the set {Heads, Tails} at random:

1. Take a coin out of your pocket and flip it.

2. Leave the coin in your pocket and talk about the side it would have landed on if it had been flipped.

Both procedures select an element from our set at random, yet only the former involves a physical process.

Of course, one could attempt to explain the conditional selection procedure by the former kind of model. One could imagine, for example, God taking a gigantic multi-sided die with all the A-worlds written on it, and rolling it at the beginning of time to determine which world is ur-selected. Even ignoring its fanciful nature, this procedure is still somewhat underdescribed: how exactly does he throw the die? Does he spin it, or throw it in the air, or use a dice tower? The random variable representing the spinning is different from the throwing, so these different procedures in principle provide competing analyses of the selection function.

Schulz [31] attempts to reduce the conditional selection process to an independent conception of arbitrary selection. According to some accounts of arbitrary reference (Breckenridge and Magidor [5]) when I say something like 'let x be an arbitrary Frenchman' then, x succeeds in referring to a particular Frenchman – Nicolas Sarkozy, prehaps. More generally there is some specific individual – the arbitrary F – for any property F whatsoever. According to Schulz's analysis, f_A is just identical to the arbitrarily selected A-world. This interpretation raises a lot of puzzles apart from questions about the underlying theory of arbitrary reference. For one thing, if I were to pick an A-world completely arbitrarily you'd have thought that it's unlikely I'd happen to pick the world that occurred if A – although according to Schulz, these two worlds are always the same. A more technical worry is that this account gets the wrong constraints on the selection function when combined with my interpretation of the selection function: in order to validate modus ponens, f_A must always be the actual world when A is true, whereas if I completely arbitrarily select an A-world I could end up picking a non-actual world even when A is true.[21]

I should emphasize, by contrast, that I don't have any reductive ambitions here: I do not think that we should expect to be able to describe some antecedently understood process of random selection – whether it be rolling a dice, or arbitrary reference – to which one can reduce the ur-selection function. In my view the selection process is closer to the model described in (2) above – it is inherently conditional, and ultimately cannot be explained in other terms. The idea that f_A is selected from A at random is essentially a helpful heuristic for motivating formal constraints on the selection function (see, especially, our discussion of CSO in sections 2.4-2.6), and for getting an intuitive picture for evaluating conditionals.

It is worth noting that if this is considered a shortcoming of the present analysis, it is one shared by the most prominent alternative. The similarity analysis of indicative conditionals along the lines of Stalnaker [34] doesn't attempt to provide a reductive account of the selection function in terms of an antecedently understood notion of similarity either. Early objections due to Tichy [42] and Fine [12] led Stalnaker to concede that if conditionals are understood in terms of an ordering of some sort, the ordering will be a backformation from our judgments about ordinary conditionals and not a notion that we had antecedently. Luckily, such an ordering can always be achieved by defining it explicitly in conditional terms: one can say that x is more 'similar' to actuality than y iff, if either x or y had obtained, x would have obtained. (Assuming some conditional logic distinctive to Stalnaker[22] this will satisfy the formal properties of an ordering.) But this ordering is clearly not one we have independently of an understanding of conditionals. The virtues of the similarity analysis therefore do not include a reduction of conditionality to an independent notion of

[21] The Lewis-Schulz interpretation avoids this consquence, and diverges from my framework, by instead defining the selection function as arbitrarily selecting from amongst the *closest*, or perhaps the most *relevant*, A-worlds. When A is true the closest A-world is unique – we are effectively randomly selecting from the singleton of the actual world. However, the introduction of similarity or relevance orderings to the analysis introduces some of the same problems I argue Stalnaker's analysis succumbs to in section 2.

[22] See the principle CSO described in section 2.4

similarity; however it does give us a useful heuristic for evaluating conditionals and motivating constraints on the selection function that provide the correct conditional logic.

This is not to say that we cannot say anything informative about the random variable f_A on our interpretation. For example, we know what the distribution of f_A is. In our informal gloss (to be made precise below) the probability that f_A is the world w is directly proportional to the probability of the world w, and the probability that f_A belongs to a subset of A is directly proportional to the probability of that subset. This property fixes an important aspect of the conceptual role of the ur-selection function that narrows down the possible candidates for f more than anything that Stalnaker says about it.

The strength of this constraint should not be underestimated. Without it one might be tempted to object that nothing I have said so far distinguishes the random world interpretation from Stalnaker's: 'the most similar A-world to actuality' is formally a random variable as well (it picks out a world at each world), so nothing in our formalism rules out Stalnaker's interpretation of f. Of course, one could lodge a similar objection to Stalnaker: if one leaves the similarity ordering completely unconstrained one could get the trivial ordering that collapses it into the material conditional, or some other competing conditional. By contrast, we will see later than any random variable generated by a similarity ordering in this way cannot have the probability distribution described above. As we shall see, the second tenet of the random world interpretation – that probability that f_A is in B is proportional to the probability of B – rules out both a similarity interpretation and the material interpretation of the selection function, whereas the formal constraints on similarity rule out my interpretation of the selection function but not the material interpretation.

So far we have been talking somewhat elliptically about 'the probability' that a given world has been selected; the kind of probability that is being invoked here is obviously central to the analysis. In our analogy with the process of selecting a ball from the bag, the selection process is naturually classified as being random if two or more balls have an objective chance of being selected. An analysis of some conditionals in terms of chances, such as Schulz's analysis of subjunctives, does indeed seem promising. However our primary concern is indicatives, and here it is much more natural to look to a subjective interpretation of probability.

Let us outline a broadly Bayesian framework in which we can formulate this idea more precisely. The Bayesian formalism assumes a special class of probability functions called the *rational ur-priors* – roughly the class of credence functions that a completely rational agent with no evidence whatsoever could possess. These functions are presumably not the credences that any ordinary person ought to have, since pretty much anything capable of having credences possesses some evidence. According to the Bayesian picture, if a person has total evidence E, then the credence they epistemically ought to have in a proposition, A, or the degree to which their evidence supports A (call this their 'evidential probability') is given by conditioning their ur-prior, Pr, on E – thus the evidential probability of A is just $Pr(A \mid E)$, the probability of $A \wedge E$ divided by the probability of E.

Clearly not every probability function can be a rational ur-prior. For example, an agent with no evidence oughtn't be able to rule out any contingent hypothesis out *a priori* – she should assign non-zero probability to contingent hypotheses.[23] But presumably there are also more conceptual constraints – elsewhere I argue that every rational ur-prior ought to assign a particular intermediate credence to the proposition that Harry is bald conditional on his having certain hair numbers in the borderline region ([REF]). Another popular constraint – the Principal Principle – states that, conditional on a hypothesis about the chance of a proposition, each ur-prior must match the hypothesised chance (Lewis [26]). It should not be too surprising, I hope, to expect that the prior probabilities of conditional propositions are likely subject to conceptual constraints like this.

The second feature of the random world interpretation corresponds, intuitively, to the idea that the more probable a world is the more likely it is to be selected. This principle, when precisified, is therefore naturally understood as a conceptual constraint on ur-priors – for any rational ur-prior, the probability that the world w is selected is proportional to the probability of w.

It is helpful here to recall our characterisation of a selection function as a collection of random variables, f_A, for each antecedent A: as we noted above, it is possible to make talk about the probability of a particular world being selected completely precise. The random world interpretion states that the probability (according to any ur-prior) that a particular A-world is selected by f_A is directly proportional to that worlds probability. If there were a finite number of worlds, this stipulation entails that the probability that a given world is selected by f_A is in fact just the probability of that world conditional on A.[24] A slightly more general thesis is needed when we are dealing with infinite spaces, since the probability of any given world is usually 0. The more general thesis states that for any (measurable) set of worlds A and $B \subseteq A$:

> Proportionality: For any rational ur-prior: the probability that the selected world f_A belongs to B, a subset of A, is proportional to the probability of B (provided the probability of A is non-zero).

Note that this entails that the probability that f_A is in B is identical to the probability of B conditional on A.[25]

There is, of course, a question about whether it is possible to have a probability space in which these random variables exist for each proposition A in the space. It turns out that if the probability space meets some very natural

[23] When the space of worlds is infinite this constraint needs to be refined somehow (see Easwaran [9]). More appropriate constraints are available if we take conditional probabilities as primitive, using the formalism of Popper functions.

[24] If $A = \{x_1, ..., x_n\}$ then the idea that the probability that $f_A = x_n$ is is proportional to the probability of x_n just means that $Pr(f_A = x_n) = \alpha.Pr(x_n)$ for some fixed constant α. The fact the $Pr(f_A \in A) = 1$ ensures that $\alpha = \frac{1}{Pr(A)}$.

[25] Proportionality says that $Pr(f_A \in B) = \alpha.Pr(B)$ for every subset B of A. In particular $Pr(f_A \in A) = 1 = \alpha.Pr(A)$ so $\alpha = 1/Pr(A)$, and $Pr(f_A \in B) = Pr(B)/Pr(A) = Pr(B \mid A)$ when $B \subseteq A$.

conditions – namely that it is an atomless *standard* probability space[26] – it will always be possible to find random variables, f_A, satisfying this condition.[27]

What is less obvious, however, is that it is possible find a single selection function (i.e. a set of random variables) that satisfies this condition relative to *every* rational ur-prior. One could maintain, with Carnap, that there is only one rational ur-prior and thus no special problem beyond that of finding random variables of the kind just described. But it would be nice to know if our theory is consistent with more permissive forms of Bayesianism in which the space of ur-priors is relatively rich. To get the relevant notion of richness on the table it helps to have a slightly more concrete model of the rational ur-priors.

Begin with a pretheoretic distinction between the categorical (non-conditional) propositions and hypothetical propositions. A crude way to test whether a proposition is hypothetical or not is whether it can be expressed by a conditional sentence; this test is not perfect, but I take it that the distinction has some pretheoretic standing so it doesn't matter too much. Categorical propositions are closed under the Boolean connectives – a negation, disjunction or conjunction of categorical propositions is also categorical – so we can assume that the categorical propositions is itself a complete Boolean algebra which is a subalgebra of the algebra of all propositions. Ignoring, for the time being, any conceptual constraints on ur-priors not directly involving conditionals, we can outline a simple condition on the space of ur-priors that guarantees that they are sufficiently 'rich'. A set of ur-priors is rich iff: for any probability function defined over the algebra of categorical propositions there is some rational ur-prior defined over all propositions, both categorical and hypothetical, that agrees with the former probability function on the probability of each categorical proposition. The resulting picture is reasonably permissive: it is rationally permissible to have pretty much any opinions about the categorical one likes, prior to recieving any evidence. On the other hand it is not so permissive as to count any (regular) probability function as a rational ur-prior: once the probabilities of the categorical propositions are fixed, the conceptual constraints involving conditionals (such as PROPORTIONALITY) do not leave much leeway in how we assign probabilities to the remaining hypothetical propositions.

[REMOVED FOR BLIND REVIEW: a brief description of some relevant results proved by the author in other work.]

1.3 Stalnaker's Thesis

At this juncture it is worth noting that our interpretation of the selection function realizes, at least in a limited form, Stalnaker's project of unifying the selection function semantics with the probabilistic theory of conditionals.

[26] A probability space $\langle W, \Sigma, Pr\rangle$ is standard if: W is a Polish space (a completely metrizable, separable topological space) and Σ is the generated Borel algebra. The probability space is also non-atomic if whenever $Pr(A) > 0$ there is a $B \subseteq A$ in Σ such that $Pr(A) > Pr(B) > 0$.

[27] Indeed, it will be possible to choose random variables, f_A, that maps each member of A to itself. This ensures that the law of modus ponens holds for the conditional. See [XXX] [REF] for a proof.

The probability that the selected world, f_A, belongs to the proposition B was defined earlier as the probability of the set $\{x \mid f_A(x) \in B\}$, which is of course the probability of the set of worlds at which the ur-conditional, $A \to B$, is true. We also saw above that proportionality entails that this probability is identical to the probability of B conditional on A. Putting this together we get:

$Pr(A \to B) = Pr(B \mid A)$ for every rational ur-prior, Pr.

which is a special case of the thesis known as Stalnaker's thesis, restricted to the ur-conditional and ur-priors.

So far we have been limiting our attention to the relation between ur-priors – credences of a completely uninformed agent – and the ur-selection function that gives the truth conditions for a conditional uttered in a context in which there is no available evidence. What of ordinary conditionals uttered by ordinary people with evidence?

Consider a piece of evidence E. A natural extension of the above equation to situations where agents have the evidence E, is the thesis that if Pr_E is the result of updating a prior on evidence E, and f_E is the result of updating a selection function on that evidence, the (updated) probability of the (updated) conditional $A \to_E B$ should be the (updated) conditional probability of B on A.

In order to do this we need to say both how to update the ur-selection function with evidence, and how to update ur-prior probabilities. The standard Bayesian answer to second question states that if my total evidence at world x at a given time is $E(x)$ then my evidential probability of A at x is $Pr(A \mid E(x)) = Pr(A \cap E(x))/Pr(A)$ where Pr represents my ur-prior. In order to update a selection function with some evidence $E(x)$, it is natural to apply Harper's constraint, outlined in section 1 (see Harper [19]): the resulting function is determined from the ur-selection function, f, by the condition $f_{A \cap E(x)}(x)$. In other words the selection function doesn't select an arbitrary A-world at random, it selects an *epistemically accessible* A-world at random, where a world is accessible at x is it is consistent with the evidence, $E(x)$, at x.

Putting this altogether we get the thesis:

CP $Pr_E(A \to_E B) = Pr_E(B \mid A)$ for every ur-prior Pr, world x and evidence E.[28]

Due to the interplay between probability function and the selection function – we update them both in tandem – this thesis is not subject to the standard triviality results. Indeed, this is a version of a contextualist response to the triviality results (van Fraassen [44], Harper [19]). I defend this particular version elsewhere (see [REF]), and show how it makes the kinds of predictions that Stalnaker's thesis was intended to make.

[28]For each world x, Pr_E is a probability function at x, given by $Pr(\cdot \mid E(x))$. Technically Pr_E is a function from worlds to probability functions – this formalism is in place to model cases where we do not know what our evidence is, and consequently do no know the actual values of the evidential probability function.

One might argue that the principle above does not deliver quite what Stalnaker's original project set out to achieve. However I think it is worth emphasising that the original project was not very well defined, and was arguably ill-conceived for reasons that had nothing to do with the triviality results. The thesis typically called 'Stalnaker's thesis' is often characterised loosely by the slogan that a rational agents confidence in a conditional proposition – the proposition expressed by the sentence 'if A then B', say – should be identical to their conditional confidence in B given A. Unfortunately this loose characterisation simply doesn't make much sense in a contextualist framework such as the one I have been describing above. To illustrate the problem, suppose that my conditional confidence in B given A is $\frac{1}{2}$. Then by the gloss of Stalnaker's thesis above my credence in the conditional *if A then B* should also be $\frac{1}{2}$. But there is no such thing as *the* conditional if A then B: there are lots of different conditional connectives that can be expressed by the words 'if' in different contexts, and consequently lots of different propositions to be expressed by the sentence 'if A then B'. More precisely, for each possible bit of evidence E, there's a corresponding conditional $A \rightarrow_E B$. Our rough gloss of Stalnaker's thesis has not told us which of these conditionals is supposed to have probability $\frac{1}{2}$.

If we were to apply the thesis flat-footedly we might take it to mean that my credence in the proposition $A \rightarrow_E B$ must be my conditional credence (i.e. $\frac{1}{2}$), for *every* possible connective $\rightarrow_E$. But this delivers completely absurd results: to take an extreme example, suppose that the evidence E corresponds to the evidence of some omniscient agent – someone who knows absolutely everything (the accessibility relation represented by the identity relation). In this case $\rightarrow_E$ just collapses into the material conditional, and the flat-footed application of Stalnaker's thesis entails that ones credence in a material conditional is always your conditional credence. With a little verification one can see that this conclusion is absurd: my confidence in the material conditional cannot always be my conditional confidence, unless my credences are completely trivial.[29]

(Note, by contrast, that an *omniscient* agent *would* have trivial credences, and would able to satisfy Stalnaker's thesis for the omniscient selection function f_E (i.e., for the material conditional). Indeed this is exactly what our more restricived principle CP says about omniscient agents: the omniscient selection function obeys Stalnaker's thesis for omniscient credences. But the flat-footed interpretation of Stalnaker's thesis entails that even ordinary people should have their credences in the material conditional match their conditional credences, and this is clearly false.)

[29]The material conditional only satisfies Stalnaker's thesis if there are no more than two propositions that I assign any credence at all.

2 Indicatives and similarity

2.1 Antecedent Strengthening

One of the most important motivations for the similarity semantics was the discovery, made very salient in places like Lewis's 'Counterfactuals' and Adams' 'The Logic of Conditionals', that natural language conditionals behave non-monotonically. A non-monotonic conditional is one that does not satisfy the law of antecedent strengthening (AS, below). Such conditionals also relinquish the closely related principles of transitivity (TR) and contraposition (CO), all of which are laws that are validated on a material or strict conditional account of indicative conditionals:[30]

AS $(A \to C) \supset (AB \to C)$

TR $((A \to B) \wedge (B \to C)) \supset (A \to C)$

CO $(A \to B) \supset (\overline{B} \to \overline{A})$.

For example, it might be reasonable to think that if Alice went to the beach, she had a good time. If one accepted AS, however, one could infer the patently absurd conclusion that if Alice went to the beach and got mauled by a shark, she had a good time. Or to adopt an example from Adams, one might well accept that if it rained yesterday, there wasn't a terrific cloudburst, but it would not be reasonable to infer by contraposition that if there was a terrific cloudburst, it didn't rain (Adams [1] p15).[31]

These observations constituted a very strong case for the similarity semantics. While all previous attempts to model conditionals had been monotonic, the similarity account provided us with a concrete model of a non-monotonic conditional. For example antecedent strengthening is not valid, since even if the closest A worlds are C worlds, it might turn out that to make B true you have to go to very disimilar worlds. The closest worlds that satisfy both A and B may not be C-worlds.

Although these observations are suggestive, and appear to favour the similarity semantics, it should be noted that it is not the only concrete model of non-monotonicity that we have on the table. Conditional probabilities behave similarly: the probability of C on A might be high while the probability of C on AB can be as low as 0 (see diagram 1 – here A is the encompassing rectangle). Observations such as these can be seen motivating theorists like Adams [1]. Note also that other interpretations of the selection function semantics, such as the random world interpretation, are also non-monotonic: a world randomly

[30] Below, and when it is convenient, I shall occasionally adopt the shorthand AB for $A \wedge B$ and $\overline{A}$ for $\neg A$. Given natural background assumptions the principles below can all be derived from one another.

[31] Note that recently some authors working within a broadly dynamic framework have challenged the moral usually drawn from this data, suggesting a monotonic strict conditional account instead (see von Fintel [46] and Gillies [15]). For further discussion, and convincing responses to these papers, see Moss [29].

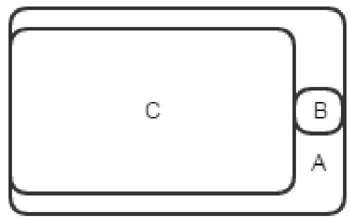

Figure 1:

selected from among the A-worlds might end up being different from one randomly selected from among the AB worlds, and consequently they might not both agree about C. Indeed, these observations are not unrelated given that the random world interpretation is consistent with a version of Stalnaker's thesis. From diagram 1 one can see that one were to select an A world at random (i.e. to pick a point in the outer rectangle) it is prettly likely that one will pick a C world, whereas if one were to pick an AB world (from the smallest rectangle) one has no chance of selecting a C world.

The idea that similarity to the actual world plays a role in determining the truth conditions of a subjunctive conditional may have an initial amount of plausibility, although there are some serious obstacles to this project concerning the relation between chance and similarity (see Hawthorne [20], Hajek [17], Williams [47], Schulz [31]). Extending the similarity semantics to indicatives is another matter altogether. Although such an extension was never part of Lewis's picture, Stalnaker [34], and more recently Nolan [30], have attempted to provide unified accounts of both indicatives and subjunctives in terms of similarity.

This project raises further hard questions, since it is not immediately obvious what respects of similarity to the actual world are relevant to the evaluation of indicative conditionals. However in my view, the primary problem is that intuitively valid principles of conditional logic place formal constraints on the indicative ur-selection function that simply cannot be met if the selection function is given a similarity based interpretation. Many of the points I am going to rehearse are extensions of familiar points in one form or another – however the reaction to these problems usually takes one of two forms. One is to weaken the conditional logic to get something that can be modeled by a similarity interpretation. Lewis's logic of subjunctives is a good example of a conditional logic that has been weakened to fit the similarity semantics, but it is not, I shall argue, a plausible logic of indicatives. Another is to drop the similarity interpretation in favour of abstract orderings of worlds (see chapter 7 of Stalnaker [39]). The motivation for this interpretation of the selection function is no longer grounded on the general heuristic of changing the world to make the antecedent true in the smallest way possible, but rather leans more heavily on the semantics ability to validate the right logic. In sections 2.1-2.3 I'll focus on the first issue, and in sections 2.4-2.6 I'll turn to the question of whether even abstract orderings will validate the right conditional logic.

2.2 Conditional Excluded Middle

Let us begin with the law of conditional excluded middle. Typically discussions of this law have centered around the subjunctive version, an instance of which is:

> Either this coin would have landed heads if it had been flipped or it would have landed tails if it had been flipped.

The status of the subjunctive version of the principle has been the subject of much discussion, and while I'm sympathetic to the principle, I think the jury is still out (see Stalnaker [37], Williams [48], Goodman [16]). However the indicative version of the principle, at least when restricted to bare indicatives, has a very different status. Imagine, now, that there is a coin on the table in the next room which might, for all you know, have just been flipped. Since we don't know whether the coin has been flipped we can't outright assert that the coin either landed heads or tails, but we can assert:

> Either the coin landed heads if it was flipped, or it landed tails if it was flipped.

To deny this version of CEM, it seems, is to take seriously the possibility that the coin was flipped and landed on its edge, or something like that. If someone were to deny this instance, then I think it would be legitimate to ask 'if neither of those options are right, how *did* it land if it was flipped?'.

The validity of CEM for bare indicatives is also born out in the way that we ask conditional questions. To answer 'neither' in the following exchange in the context described above sounds unnatural. A response of 'I don't know' seems more appropriate:

Q Which way do you think the coin landed if it was flipped: heads or tails?

A # Neither

A I don't know.

The validity of conditional excluded middle for bare indicatives strikes me as a piece of data that any account of indicatives ought to be able to accommodate, not a controversial principle like its subjunctive cousin.[32]

A more theoretical argument for conditional excluded middle can be given from the perspective of views, such as the one outlined in section 1, that validate

[32] One could attempt to explain why 'Neither' is a marked by giving a fleshed out account of the presuppositions of conditional questions. But such an explanation would also fail to predict the appropriateness of the seemingly correct response 'I don't know'. One cannot in general respond to a question with a failed presupposition with 'I don't know' – to answer 'I don't know' to 'Has Alice quit smoking?', for example, one acknowledges the presupposition that Alice was drinking at some point.

the intuition that the probability of an indicative conditional is the same as the probability of the consequent conditional on the antecedent. Accepting this intuition, at least provisionally, we can see that the probability of conditional excluded middle must always be 1. The reason for this is that $Pr((A \to B) \vee (A \to \neg B) = Pr(A \to B) + Pr(A \to \neg B) - Pr((A \to B) \wedge (A \to \neg B))$ by probability theory. Assuming, quite plausibly, the validity of the principle of conditional conglomeration, the probability of $(A \to B) \wedge (A \to \neg B)$ is less than or equal to the probability of $A \to (B \wedge \neg B)$, which is $Pr(B \wedge \neg B \mid A) = 0$. Thus the right hand side becomes $Pr(B \mid A) + Pr(\neg B \mid A) = 1$ by the conditional to conditional probability link. (Of course, Stalnaker's thesis can't hold in full generality, but in the random world interpretation it holds for a wide enough range of probability functions for this to argument to be compelling: the probability of CEM for the ur-conditional, for example, must be 1 relative to every rational ur-prior.)

Note of course that there are conditionals that are in the indicative mood, ones that I set aside earlier, that do not behave like this. Consider for example:

> Either this coin will land heads if it is flipped or it will land tails if it is flipped.
>
> Either the coin lands heads if it is flipped, or it lands tails.

Both of these instances strike me as controversial in the same way as the subjunctive version are.

The first of these is not a bare indicative conditional: it contains the modal auxiliary 'will' in the consequent. Such conditionals are sometimes just alleged to be subjunctives in a different tense (Dudman [8]), in which case it is hardly surprising to find them patterning with the subjunctive versions of CEM. But more importantly, even if this sentence were indeed false, it would not be a counterexample to the principle we are interested in. It stands to the bare version of conditional excluded middle in the same way that the sentence 'either there will be a sea battle tomorrow or there will not be a sea battle' stands to the bare version of excluded middle. The sea battle example is also arguably contentious, but it has little bearing on the ordinary law of excluded middle, since the negation is in the scope of 'will'. The latter example seems to have the logical form $\mathsf{Will}A \vee \mathsf{Will}\neg A$ and the former $(A \to \mathsf{Will}B) \vee (A \to \mathsf{Will}\neg B)$.[33] Whatever you think of the future directed versions of these principles they are not straightforwardly instances of excluded middle or conditional excluded middle as I have defined them.

The second conditional also doesn't appear to be a real instance of conditional excluded middle either, since it seems to have a covert habitual operator: despite its surface form, the fundamental structure of the conditional is something like 'the coin usually lands heads if it is flipped, or it usually lands tails',

[33] As mentioned earlier, some theorists have taken the linguistic data surrounding the interaction of modals, such as 'will', and conditionals to indicate that conditionals are not binary connectives taking a modalised sentence in the consequent. However see Dorr and Hawthorne [7] for explanations of these phenomena in a more orthodox setting in which conditionals are represented by connectives.

or perhaps 'the coin is disposed to land heads if it is flipped or it is disposed to land tails if it is flipped'. This is consonent with the orthodox theory of the simple present tense in which simple present tensed sentences contain covert generics at the logical form. To demonstrate this, note that there are analogous puzzles to be made about simple present tensed sentences that aren't conditionals. Suppose Alice sings in her local choir, Bob doesn't sing as a matter of principle, and Caroline sings infrequently. It seems OK to say 'Alice sings' and 'Bob doesn't sing', whereas 'Caroline sings' and 'Caroline doesn't sing' both seem to express falsehoods. According to the covert operator theory, the law of excluded middle is not in contention here because these sentences have hidden structure.[34]

The validity of the law of conditional excluded middle imposes some constraints on the indicative selection function. In the generalised Chellas semantics, $f(A, x)$, must output exactly one world for any consistent antecedent A. For suppose that $f(A, x)$ was a set of worlds with two or more members. Then we can always find a proposition, B, that overlaps the set $f(A, x)$ and whose negation also overlaps $f(A, x)$. The selected A worlds, $f(A, x)$, are neither all B worlds nor are they all $\neg B$ worlds, whence a failure of CEM is generated. This is, of course, just abstract model theory, but it gives us a feel for what the correct interpretation of the selection function must look like: there is a unique epistemically possible state of affairs that is the way things are if A.

Now, even if the similarity interpretation of the subjunctive selection function is correct, there is, I think, a special problem here for an analogous treatment of indicatives. For if $f(A, x)$ represents the set of most similar accessible worlds to x at which A is true, then there ought to be plenty of instances in which $f(A, x)$ contains multiple worlds: whenever there are several equally similar accessible A-worlds that are more similar to x than any other accessible A-worlds. In the dispute about subjunctives, it is tempting to take this observation as a tiebreaker in favour of the CEM deniers. But in the indicative case, it is not plausible to lay the blame at the feet of CEM – I think it is much more natural to take this problem as indicating a problem with the similarity interpretation of the indicative selection function.

One rather hamfisted way to solve this problem, whilst keeping within the spirit of the similarity interpretation, is to take something like the pretheoretic ordering of similarity and flatten it out in some arbitrary way: when we come across two worlds that are equally similar to the actual world, we break the tie arbitrarily to create a total ordering of worlds. Perhaps we can also consider all possible ways of flattening out the similarity relation and supervaluate over them (Stalnaker [37]). On this proposal, the relation between conditionals

[34] An exception to this rule of applying a covert habitual modifier occurs when one narrates events in real time using the present tense: for instance, 'He runs, he shoots, he scores!' as uttered by a football commentator. In these cases the law conditional excluded middle seems to be valid once again. Suppose, for example, that John always flips a coin to determine whether he has poached or boiled eggs if it's a Tuesday. We can narrate John's daily routine the simple present: John wakes up, has a coffee, and either has boiled eggs if it's Tuesday, or has poached eggs if it's Tuesday. If this is right, it provides further confirmation of the hypothesis that the bare versions of conditional excluded middle are valid.

and similarity is slightly more complicated and indirect. There are two different kinds of orderings: the similarity ordering orders worlds by how similar they are to actuality, and then there is an abstract ordering (or a collection of abstract orderings) whose function is given by the role they play in assigning truth conditions to conditionals. An abstract ordering might rank y before z even if y is in fact no more similar to actuality than z. The connection between conditionals and similarity is given by the constraint that the abstract ordering *respects* the similarity ordering: if y is more similar to actuality than z then y must be ranked before z by the abstract ordering.

One of the toted benefits of the heuristic of similarity was its ability to provide intuitive explanations for why certain inferences were valid or invalid. Yet in the case of CEM the heuristic does not predict its validity – it must be put in by brute force. In this regard, the random world interpretation gives a much more elegant and straightforward explanation for the validity of conditional excluded middle. The ur-selected A-world is one of the A-worlds, but it could be *any* of the A-worlds whether similar to actuality not.

2.3 The Limit Assumption

The flattening out approach does not solve all our problems. Consider the following infinitary principle:

LIM $A \to B_1, A \to B_2, ... \vdash A \to \bigwedge_n B_n$

LIM and CEM are independent of one another – one can accept one without the other.[35]

This principle roughly says that the conjunction of things that are true if A is also true if A – if B_1 holds if A, B_2 holds if A and so on, then B_1, B_2 and so on all hold together if A. As Kit Fine points out in [13], this principle has a great amount of intuitive appeal. The finitary version of it is a part of almost every logic of conditionals on the market, but Fine notes it is hard to motivate the finitary version without also motivating the infinitary version: 'what makes the finitary rule plausible is the more general principle that the logical consequences of the counterfactual consequences of a counterfactual supposition should also be counterfactual consequences of the supposition. But if this is the justification of the finitary rule, then it serves equally well to justify the infinitary rule.' (Fine is of course talking about the subjunctive case here, but the reasoning is no less compelling in the indicative case.)

This principle is not validated in Lewis's semantics for counterfactuals. According to Lewis, $A \to B$ is true if there's an A world such that every A-world as or more similar to the actual world than it is also an B-world. There could be an infinite descending chain of ever more similar A-worlds such that for any

[35]To construct a model of CEM without LIM is non-obvious. The trick is to construct a non-standard selection function that, instead of outputting a world, outputs a non-principal ultrafilter over the set of worlds. The conditional 'if A then B' is true at x if the set of worlds B belongs to the ultrafilter $f(A, x)$.

i, B_i is true from some point onward in that sequence, but there is no point at which every B_i is true simultaneously.

This observation is not, however, an argument against the principle LIM since it relies on a particular semantics whose appropriateness rests on how it treats principles of conditional logic, including LIM itself. In fact Lewis's semantics does particularly badly in this regard: according to Lewis's semantics it follows that, for *any* number $\epsilon > 0$, if it had been hotter than 85, it would have been cooler than $85 + \epsilon$ degrees (assuming it is in fact cooler than 85 degrees). Of course, there is no temperature strictly between 85 and $85 + \epsilon$ for every $\epsilon > 0$, so the collection of things that would have obtained if it had been hotter than 85 degrees is jointly inconsistent (see Herzberger [21], Fine [13]). Call this kind of situation – where A is consistent but the set of things true if A isn't – a 'conditional inconsistency'.

If one accepts conditional excluded middle, as I think one should for bare indicatives, then conditional inconsistencies are more than just a peculiarity of Lewis's particular way of invalidating LIM. Any failure of LIM whatsoever will involve a conditional inconsistency. Suppose that LIM fails: there is a situation in which for each n, $A \to B_n$ holds, and $\neg(A \to \bigwedge_n B_n)$ also holds. The negated conditional entails $A \to \neg \bigwedge_n B_n$ by CEM. This means that the total collection of things true if A is joint inconsistent, despite the fact that A is consistent (in the sense that $\neg(A \to \bot)$). We know that the sequence of propositions $B_1, B_2, \ldots$ plus the negated conjunction $\neg \bigwedge_n B_n$ are jointly inconsistent, but also just saw they each of these things are true if A is true.[36]

As with CEM, there is also a more theoretical argument for LIM deriving, again, from the intuition that a conditional's probability is the probability of the consequent given the antecedent. Given countable additivity, if the probability $Pr(B_n \mid A)$ is 1 for each n then so is the probability of $Pr(\bigwedge_n B_n \mid A)$, which suggests that anyone who accepts all the premises of LIM and ought to accept the conclusion provided they assign their credences in conditionals according to its conditional credence.

At any rate, given (i) the apparent validity of CEM for indicatives and (ii) the implausibility of conditional inconsistencies, we have good reasons to think that LIM is valid for indicatives. But once again this poses a problem for the similarity account of indicatives, for as we saw earlier LIM can fail whenever there are infinite descending chains of worlds of increasing similarity to the actual world. It is surely true, according to any reasonable notion of similarity, that there can be infinite sequences of worlds of increasing similarity. Note, for example, that if there are no infinite descending chains then there have to be infinite *ascending* chains of worlds of decreasing similarity: the only way to avoid having chains of either kind is for there to be only finitely many worlds in total! But given that there can be infinite ascending chains, it is puzzling why there is this stark asymmetry.

[36] Without CEM one cannot always generate conditional inconsistencies from failures of LIM. For example, imagine that W is an infinite set of worlds: one could let $A \to B$ be true iff either A and B are true, or A is false and B is cofinite.

Indeed, one could imagine two worlds x and y, such that (a) from xs perspective y is separated from it by an infinite ascending chain of worlds of decreasing similarity, and (b) the similarity facts are the same at both x and y: x and y both agree about how similar any other two worlds are. It follows that from ys perspective, there must be an infinite *descending* chain of increasingly similar worlds separating y from x, so (a) and (b) cannot be jointly satisfied. The only way to maintain a similarity type of theory, then, is if the similarity measure is radically contingent: only worlds seperated from x by a finite number of worlds (according to xs similarity measure) can agree with x about the similarity facts. There are never more than countably many worlds seperating from x by finitely many worlds, so there will have to be plenty of contingency regarding what makes two worlds similar.

Note also that the strategy of 'flattening out' an ordinary similarity ordering will not always result in an ordering without infinite descending chains. One might hope for an operation analogous to our 'flattening out' operation for getting rid of ties: something that takes a similarity ordering that has infinite descending chains as input, and gets rid of them somehow outputting a relation without infinite descending chains. Unfortunately, there is no plausible way of doing this. Suppose R is the result of performing some operation to a similarity ordering with infinite descending chains, and R has no infinite descending chains: then it follows that for some x and y, x is ranked closer to actuality than y according to R even though y is in fact more similar to actuality than x. In fact, x will be ranked closer by R than infinitely many worlds that are in fact more similar to actuality than x. (Reason: suppose that X is a set of worlds such that for every world in X there's another world in X that is more similar to actuality. Then the minimal element of X according to R, call it x, will be less similar to actuality than each member of the infinite subset of X consisting of worlds more similar to actuality than x.)

A final point about all this, is that it is often assumed that an ordering of possible worlds by their similarity to a given world, x, is a result of a more general quantitative structure that tells us, for any pair of worlds, how similar they are to each other by giving us a *degree* of similarity (this intuition is pervasive – for a recent example, see Kment [22]). Structures of this kind are called metric spaces, and are given by a distance function, d, that outputs a positive real number, $d(x, y)$, for any pair of worlds x and y, telling you how close they are.[37] As we've shown already, similarity facts have to be contingent on this picture so in principle each world, x, will come with its own similarity metric, $d_x(y, z)$. Now one can attempt to define a selection function by letting the value of $f(A, x)$ be the world y in A whose distance from x according to d_x is minimal. This is not well defined for all metric spaces because there might be infinite descending chains or multiple worlds that are equally close to x. Our discussion so far effectively amounts to a constraint on our similarity metric, d_x,

[37]This function will be subject to certain constraints: the distance between x and y is 0 only if $x = y$, the distance between x and y and between y and x is always the same, and the sum of the distances between x and y and between y and z is always greater than or equal to the distance between x and z.

at a world x: it must be a metric in which for any set A there's always a unique y in A whose distance from x is minimal. Unfortunately if the set of worlds is uncountable, which it surely is[38], there is no metric meeting this constraint at any world.[39] Similarity orderings are orderings of similarity only by name – they are not capturing comparisons of distance from reality by any ordinary notion of distance.

The structure of all of these arguments is similar. One assumes the validity of LIM and CEM and combines that with the idea that the truth conditions of a conditional can be specified by abstract orderings of some kind. The conclusion in each of our arguments was that in order to validate the correct conditional logic the abstract orderings do not correspond to a measure of similarity in any ordinary sense. The usual motivation for adopting an ordering semantics is that the abstract orderings are supposed to represent a measure of similarity to the actual world. In the case of Stalnaker [34], the abstract orderings are not supposed to be orderings of similarity themselves, but are still supposed to be related in some important way to an ordering of similarity.

We can make this a little more precise by invoking a very general kind of ordering semantics where each world x is equipped with an ordering $\preceq_x$, but where the orderings themselves do not have any formal constraints built into them other than the requirement that they be orderings: i.e. that they be transitive and reflexive. On this semantics we can give a clause for the conditional:

> Abstract Ordering Semantics: $A \to B$ is true at x if and only if either there are no accessible A worlds, or for some world $y \in A$ and every $z \in A$ with $z \preceq_x y$, $z \in B$.

This theory will deliver Lewis's theory and Stalnaker's theory as special cases, depending on which further constraints on the orderings are imposed. I shall also require generally that $x \preceq_x y$ for all y, which guarantees the validity of modus ponens. The informal arguments we gave above correspond to the following theorems:

> Suppose that LIM and CEM are valid in some model based on abstract orderings over the set of possible worlds.[40] Then it follows that:
>
> (i) If the ordering of worlds by some notion of similarity to x contains a

[38]We can partition spacetime into countably many planet sized regions, and presumably for any collection of these regions it could have been the case that exactly these regions contain planets and no others.

[39]Suppose that W is uncountable. Provided f is defined on every non-empty set, we can inductively generate an uncountable sequence of worlds $x_\alpha = f(W \setminus \{x_\beta \mid \beta < \alpha\}, x)$, and an uncountable strictly increasing sequence of real numbers $r_\alpha = d(x, x_\alpha)$ (it must be strictly increasing because of the way the selection function was defined). This is a contradiction: there are no uncountable strictly increasing sequences of reals, because between any two reals there is a rational number so any mapping g such that $g(\alpha) \in (r_\alpha, r_{\alpha+1}) \cap \mathbb{Q}$ would constitute an injection from an uncountable set to the rationals which is impossible.

[40]Here I do not mean the set of possible worlds of some model, which may be any abstract set of entities, I mean the *real* possible worlds.

tie, then for some y and z, $\preceq_x$ ranks y before z even though y is not more similar to x than z.

(ii) If the ordering of worlds by some notion of similarity to x contains an infinite descending chain $\ldots y_2, y_1, y_0$, then for some n there are infinitely many worlds strictly more similar to x than y_n that $\preceq_x$ ranks as strictly less similar to x than y_n.

(iii) If x and y are separated by an infinite $\preceq_x$-chain, then $\preceq_x$ and $\prec_y$ disagree about the relative ordering of infinitely many worlds.

(iv) Either $\preceq_x$ is not representably by a similarity metric for any world x or there are only countably many worlds. (The ordering, $\preceq_x$, is representable by a similarity metric if and only if there is some metric d such that $y \preceq_x z$ iff $d(x, y) \leq d(x, z)$).

Our options therefore are either to give up LIM or CEM, reject the abstract ordering semantics altogether, or accept the abstract ordering semantics but sever the tie between the orderings and similarity in a fairly radical way. Lewis pursued the first strategy in the case of subjunctives but, as we have argued above, it is a harder sell in the case of indicatives. The last strategy seems unpalatable given that the standard motivation for the abstract semantics was its interpretation in terms of similarity. Stalnaker has attempted to partially sever the tie between the abstract orderings and similarity: while the orderings don't directly represent a similarity ordering, they are supposed to be constrained by similarity in an important way. The results (i), (ii) and (iv) above, however, place limits on how much mileage we can really get from this approach.

It should of course be acknowledged that it has been known for a long time that the similarity ordering relevant for evaluating conditionals is not a measure of similarity in any pretheoretic sense; this fact came out in early reviews of Lewis's 'Counterfactuals'.[41] The similarity ordering has to be doctored to accord with our intuitions about conditionals. Even so, the above considerations seem to show that if we are to get the right logic, the relevant orderings are formally unlike similarity orderings, whether doctored or not. They must be total orders and well founded, and are not representable by a metric of similarity unless there are a small number of worlds. The best option, in my view, is to reject the analysis of indicatives in terms of orderings altogether. But let us now set aside the question of whether such a semantics can really be underlined by a coherent notion of similarity and see if we can find some independent reasons to reject even a semantics based on abstract orderings of some sort.

2.4 The logic of indicatives: CSO

Stalnaker's formal logic of conditionals is obtained by placing one further constraint on the selection function, in addition to the constraints ID and MP from section 1:

[41] See Fine, [12], Tichy [42].

CSO If $f(A, x) \in B$ and $f(B, x) \in A$ then $f(A, x) = f(B, x)$.

Reflection on Stalnaker's interpretation of the selection function quickly reveals why this constraint must be imposed. If the closest A-world is a B-world, and the closest B-world is an A-world, then by elementary properties of closeness, the closest A-world and the closest B-world must be one and the same.[42] A selection function that does not meet this constraint cannot be represented by any kind of ordering over worlds: there will be no ordering of worlds, whether it corresponds to an intuitive ordering of similarity or not, such that $f(A, x)$ can be identified with the minimal A-world according to that ordering.

By contrast, the principle CSO is not valid on the random world interpretation of the selection function. If there are two or more worlds in the region where A and B overlap, and I were to select an A-world and a B-world at random there's a chance that I select two different worlds in their intersection. In which case the selected A-world is a B world, and the selected B-world is an A-world, but they're not the same world. The principle CSO seems to be a primary point of contention between the two interpretations.

The logical principle that this constraint corresponds to is the following:[43]

CSO $(A \leftrightarrow B) \supset ((A \rightarrow C) \supset (B \rightarrow C))$.

Roughly, if A and B conditionally imply one another, they conditionally imply the same things. The truth of $A \leftrightarrow B$, by the CSO constraint, implies that the selected A-world and the selected B-world are identical, and thus that the former is a C-world only if the latter is in virtue of being identical. The validity of CSO is much contested, and there is a small but not insignificant literature proposing counterexamples to this principle as it applies to subjunctive conditionals (see Tichỳ [41], and the discussion in chapter 7 of Stalnaker [39], Maartenson [28], Tooley [43], Ahmed [2] and Fine [13].[44])

In my view the matter is far from obvious; the principle is complicated enough that it is hard to get concrete judgments. However, I do think there

[42] This principle generalises to the Chellas semantics: if all the closest A-worlds are B-worlds and conversely, then the closest A-worlds must be the same as the closest B-worlds.

[43] There are a collection of principles that are also often discussed in connection with the similarity semantics. I list some of them below:

CV $\neg(\phi \rightarrow \neg\psi) \supset ((\phi \rightarrow \chi) \supset (\phi \wedge \psi \rightarrow \chi))$

RCV $(\phi \rightarrow \psi) \supset ((\phi \rightarrow \chi) \supset (\phi \wedge \psi \rightarrow \chi))$

RT $(\phi \rightarrow \psi) \supset ((\phi \wedge \psi \rightarrow \chi) \supset (\phi \rightarrow \chi))$

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

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

If CEM is not present, these principles bear interesting logical relations to one another. However in the portion of the logic that both the random world and the ordering semantics share (or at least, Stalnaker's version of the ordering semantics), these principles are all logically equivalent to CSO. Thus I can without loss of generality restrict my attention to CSO.

[44] Fine regards his counterexample as showing that subjunctives are hyperintensional, rather than as regarding it as a counterexample to CSO.

are a number of indirect considerations which taken together form a reasonable case for the invalidity of the principle.[45]

That said, even according to the interpretation of the selection function I prefer, CSO has a number interesting probabilistic features, some of which might even explain the initial appearance of validity. Assume the version of Stalnaker's thesis suggested in section 1.3: that in most contexts the probability of a conditional is the conditional probability. This constraint sheds some light on questions concerning which principles of conditional logic are valid: some principles must have a probability of 1 – such as the law ID ($A \to A$) or CEM – whereas other principles can have a probability strictly less than 1. If a principle can recieve a probability less than one, this suggests that it is not valid. However, there is a remarkable distinction to be drawn between two different types of invalid principle. Some invalid principles – such as the material conditional law $\overline{A} \supset (A \to B)$ and antecedent strengthing – can have arbitrarily low probability; instances whose probability is as close to 0 as you like. However there are some invalid principles whose probability cannot sink below $\frac{1}{2}$. Surprisingly CSO – indeed, any theorem of Stalnaker's logic – has this property.[46] This suggests that even if we do not accept the similarity theory, Stalnaker's logic is of interest for a different reason: the logic of the random world interpretation is the logic of things having probability 1, whereas Stalnaker's logic is the logic of things having probability greater than $\frac{1}{2}$.[47]

At any rate, if we are looking for a counterexample to CSO – a scenario where an instance of CSO seems obviously false – this observation suggests we are not going to find one. When we are looking for a counterexample we tend to try to imagine a scenario in which $A \leftrightarrow B$ and $A \to C$ are evidently true, and in which $B \to C$ is evidently false – such scenarios will be hard to come by, for this would require one to assign 0 credence to an instance of CSO, which

[45] One putative consideration in its favour, which I think can be discounted immediately, is the observation that CSO follows from the principle of transitivity for the conditional: from $A \to B$ and $B \to C$ infer $A \to C$. Although transitivity may look appealing, it has intuitive counterexamples and entails the principle of antecedent strengthening which we rejected earlier (it's a logical truth that $A \wedge C \to A$, so from the premise $A \to B$ we can always infer $A \wedge C \to B$ by transitivity).

[46] The easiest way to see this is to show that the law $(A \to BC) \supset (AB \to C)$ (an equivalent of CSO in Stalnaker's system) cannot sink below probability $\frac{1}{2}$. The only way this could happen would be if the antecedent $A \to BC$ had probability above $\frac{1}{2}$ and the consequent had probability below $\frac{1}{2}$. But if BC takes up more than a half of A, then C must take up more than half of AB (formally note that $Pr(BC \mid A) = Pr(ABC)/Pr(A) \leq Pr(ABC)/Pr(AB) = Pr(C \mid AB)$ since $Pr(AB) \leq Pr(A)$). (It should be noted that the law $(A \to BC) \supset (AB \to C)$ has the same probability as CSO if we assume: (i) modus ponens, (ii) conditional agglomeration (that $(A \to B) \wedge (A \to C)$ is equivalent to $A \to BC$)), and (iii) Stalnaker's thesis. Stalnaker's thesis guarantees that some other bits of conditional logic needed in the proof of this equivalence have probability 1.)

[47] Although I have not yet been able to prove this yet, I conjecture more that the strongest logic of conditionals whose theorems all have probability 1 is the logic I prefer, and that Stalnaker's logic is the strongest logic of conditionals whose theorems all must have probability great than $\frac{1}{2}$.

we've just argued to be impossible.[48] Other than the fact that it is quite hard to evaluate, this, I think, explains why CSO seems quite resistant to straightforward counterexamples.

A common justification for CSO is not based on direct intuitions at all, but based on fairly theoretical considerations: an inference to the best explanation – CSO is a consequence of a more general and predictive theory of conditionals based on similarity. The similarity theory, for example, gives the correct verdicts with respect to antecedent strengthening and the other non-monotonic behaviour that needed explaining, and a commitment to CSO is just a consequence of accepting these kinds of explanations.

Insofar as this kind of reasoning is motivating CSO, it is worth reiterating that CSO fails according our alternative heuristic for evaluating $f(A, x)$, in which it represents an A world that has been selected at random.

Another issue with this kind of justification is that the similarity based theory is predictive only if we already know how to make comparisons of similarity prior to evaluating the conditionals themselves. Observations in the last section and elsewhere undercut this initial motivation for accepting CSO as a formal constraint on the selection function; if we can represent the selection function in terms of some ordering at all, the notion of similarity will be a backformation from our judgments about conditionals, and not the other way around. It is for these kinds of reasons Stalnaker talks in later work of selection functions as being generated from 'abstract orderings' (see e.g. [39]), and motivates a selection function semantics based on these abstract orderings instead on the grounds that doing so validates the correct logic. To properly evaluate the case for CSO we have to evaluate its plausibility directly.

2.5 Against CSO: the dartboard

Let us turn to an example that highlights the power of CSO, and also illustrates the difference between the similarity semantics and the random world semantics. Imagine that there is a dartboard in the next room with two circular overlapping regions, A and B, drawn on it. The region at which A and B overlap is furthermore divided into two regions, U and L ('upper' and 'lower'; see diagram 2). A dart has been thrown at the board, but we do not have any idea where the dart landed. Now suppose I present the following two (seemingly compatible) hypotheses for consideration:

1. If the dart landed in A it landed in U.
2. If the dart landed in B it landed in L.

I take it that both of these conditionals are true for all we know. In the first case, presumably either the dart landed in U if it landed in A, or it landed in

[48] These observations are closely related to the fact that CSO is equivalent a rule of proof that is probabilistically valid it the sense of Adams [1]: $A \to B, B \to A, A \to C \vdash B \to C$. Roughly if one has high conditional confidence in B on A, A on B and C on A, then one should have high conditional confidence in C on B.

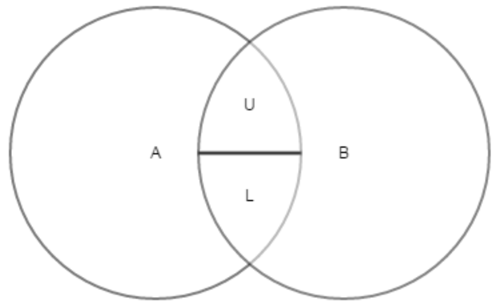

Figure 2:

L if it landed in A, or it landed in $A \setminus B$ if it landed in A, and we don't know which of these possibilities obtains. A similar disjunction is true concerning what happened if the dart landed in B (these disjunctions can in fact be proved from CEM given some modest background logic).

Now both the similarity theory and the random world theory correctly predict in this situation that both 1 and 2 are true for all we know. However it is natural to accept the stronger possibility that that both might be true together. For all we know the dart landed in U if it landed in A, but landed in L if it landed in B. This intuition is vindicated in the random selection model: for if I select an A-world at random and independently select a B-world at random, there's a chance that the first world will be in U and the second in L.

On the other hand, any theory of conditionals that validates the principle CSO predicts that 1 and 2 are in fact inconsistent. Note that 1 entails that if the dart landed in A it landed in B (because U is a subset of B). Similarly, 2 entails that if the dart landed in B it landed in A. So we have $A \leftrightarrow B$, and we have $A \to U$ by 1. CSO would allow us to infer $B \to U$, which seems to contradict 2: the only way for $B \to U$ and $B \to L$ to both be true is if conditionals with B as an antecedent are vacuously true. Presumably this isn't the case here because it's epistemically possible that the dart landed in B our example.

If you, like me, have the intuition that 1 and 2 are compatible with one another then this is something the random world interpretation of the selection function can accommodate which the similarity intepretation cannot. While I don't want to rest too much on any particular example (the consistency of 1 with 2 is admittedly hard to evaluate), I do think a pattern of judgments emerge that when all taken together tend to favour the random world interpretation over the similarity based interpretation.

Let's consider another example involving dartboards.[49] This time imagine that there is only one closed circular region, A, and suppose also that it is quite small relative to the total area of the dartboard (see diagram 3). As before, a

[49] A structurally similar example to the following is also discussed in Edgington [10] §4.1. Edgington attributes the example, used in a slightly different context, to James Studd.

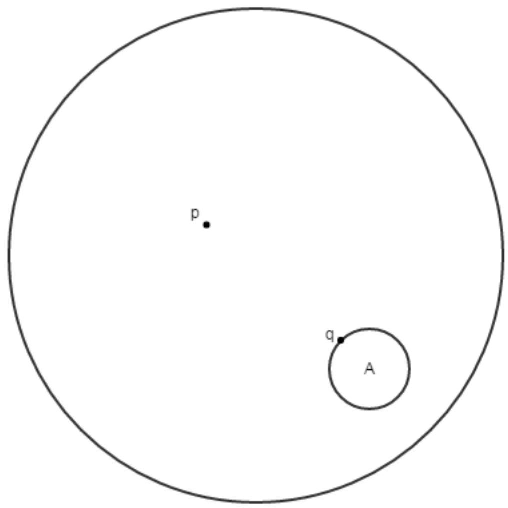

Figure 3:

dart has been thrown at the board but we do not know where it has landed: we shall simply assume that the probability of the dart landing in any subregion of the dartboard is just proportional to the area of that region. To model this scenario using the similarity semantics we need to determine some kind of way of measuring similarity between possible worlds. Since the only facts that can vary that are relevant in this example are facts about where the dart landed, it is extremely natural to build a model in which the worlds are simply identified with points on the dartboard. Each point on the dartboard simply represents a world where the dart landed on that point.

A particularly simple way to measure the similarity of two worlds is by the distance between where the dart lands in each of those worlds. I shall generalise these considerations to arbitrary orderings of similarity in a moment, but it is worth noting that it would be *prima facie* surprising if some other measure of similarity between landing places were operative in this example. Observe also that because of the shape of A there is always a unique point in A closest to any point on the dart board: in our diagram, for example, we have marked the point p and the closest point in A to p, q.[50] The problems involving uniqueness and the limit assumption from the last section can be set aside, at least in this special case.

Now consider the following conditional:

If the dart landed in A, it landed on the edge of A.

[50]The reason we can always assume a closest point in A is because A is both convex and closed.

Intuitively you should be almost certain that this conditional is false. The edge of A is just an infinitely thin line with no area, and so it is incredibly unlikely that the dart landed on the edge of A if it landed in A. Although it's not impossible the probability is presumably 0.

Now compare the verdicts we get from the two interpretations of the selection function. If we take a point p inside A, the closest point to p in A is clearly itself. If you take any point p outside A, the closest point to p in A will lie on the edge of A. So according to the similarity analysis of indicatives, the above conditional is true at all of the worlds where the dart lands outside or on the edge of A, and is false at worlds where the dart lands strictly inside A. So the probability of this conditional is simply the probability that the dart doesn't land in A. Since A takes up a relatively small proportion of the dartboard, the probability that dart doesn't land in A is high. Thus according to this similarity analysis we should be pretty sure that the dart landed on the edge if it landed in A.[51]

If we apply the random world heuristic, however, you get the intuitively correct verdict. Suppose that p is a point outside A. To see if this conditional is true you don't pick the closest point in A to p – which will definitely be on the edge if the dart lands outside A – you pick a point from A at random, which will almost certainly not be on the edge. The probability that you'd pick a point on the edge of A, if you select randomly from A without any preference for points closer to where the dart actually landed, is 0.

This is, I think, a *prima facie* consideration in favour of the random world heuristic. However there is a significant amount of wiggle room for the similarity theorist here, since this argument relied on the assumption that the more the darts landing place differs from its actual landing place, the less similar to actuality it is. Although it might seem like, other things being equal, the closer the dart lands to its actual position the more similar it is to actuality, theories of this kind are already in the business of denying our pretheoretic intuitions about the relevant notion of similarity. To avoid the above result one might try to come up with an ordering of worlds in which similarity is not measured purely by the distance between the points at which the dart lands which also gives a more sensible probability to the above conditional.

Unfortunately, this move doesn't really help us: we can generate a 'revenge' version of this problem. We saw above that measuring similarity by distance from the actual landing place gets the wrong results for the particular region A. However, it turns out that whatever similarity measure between points you choose, it's always possible to construct a region like A which gives whacky results. That is, although it is true that for any particular region of the dartboard,

[51] Indeed a similar problem applies to the Lewis-Schulz version of the theory, which lies somewhere between my version and the similarity theory. Recall that according to that theory a conditional is true at a world if the consequent is true at a randomly selected closest antecedent world. However, the present example has been constructed in a way so that all the closest worlds to a world where the dart lands outside A are worlds where the dart lands on the edge (roughly, because A is convex). So this theory similarly delivers the result that one should be almost certain that the dart landed on the edge if it landed in A.

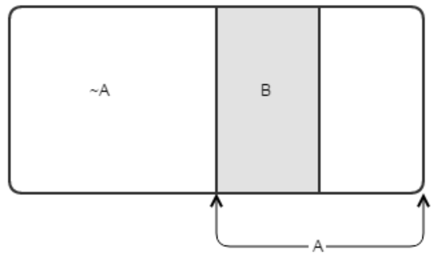

Figure 4:

A, we can choose an ordering that gives the claim 'if the dart landed in A it landed in B' the intuitively correct probability, it is not possible to construct an ordering that does this for every region at once.

To see this, suppose instead that we consider a slightly different dartboard – the rectangular one depicted in figure 3 4. Although less dramatic, we can get the same worry as before here: it seems like it should be fifty-fifty whether the dart landed in B if it landed in A. As before, it would be disasterous if we measured similarity by distance between landing places on the dartboard: we'd get that it was quite likely (75%) that the dart landed in B if it landed in A, since the conditional 'if the landed in A it landed in B' is true at any world where the dart in fact didn't land in A (if it lands in $\overline{A}$ the most similar (as measured by distance) A worlds are always B worlds).

Suppose we relax the constraint that the degree of similarity between two landing spots is simply measured by the distance between those spots. Perhaps we could find a measure of similarity according to which roughly half of the points in the $\overline{A}$ region regard a point on B to be the most similar point in A, half which don't.[52] (For full generality we must assume that each world has its own measure of similarity, and each world x determines an ordering of worlds telling us how similar they are to x according to xs similarity measure. Our argument won't turn on this.)

Now we can get the revenge problem going. The idea is to expand A to a larger set, A^+, by adding to it the set X, where X is the set of of points on the left half of the dartboard (i.e. points in $\overline{A}$) which regard a point in B to be the most similar point to it lying in A. In other words:

> A^+ is the set of points on the dartboard given by $A \cup \{p \in \overline{A} \mid$ the 'closest' point to p in A belongs to $B\} = A \cup X$

Remembering, again, that 'closest' here does not necessarily mean closest in terms of distance. Now notice that according to the similarity theorist the conditional 'if the dart landed in A it landed in B' is the set of worlds where the

[52]Indeed, formally it is always possible to find *some* ordering that has this feature, if one assumes the axiom of choice.

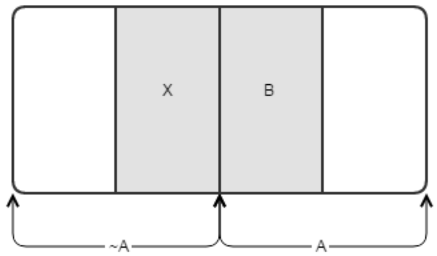

Figure 5:

most similar A-world is a B-world. If w is an A-world, then clearly the most similar A-world to w is a B-world if and only if w is a B-world. If w is not an A world, then by definition the most similar A-world to w is a B-world if and only if w belongs to X. Thus the conditional 'if A then B' denotes the set $B \cup X$.

This gives us a handle on how big X is. Intuitively it is fifty-fifty whether the dart landed in B if it landed in A. So the probability of $B \cup X$ is a half. Since we know B takes up half of A, that means that X must take up half of $\overline{A}$, as in the diagram 5. But this means, as one can see from the diagram, that $A^+ = A \cup X$ takes up $\frac{3}{4}$s of the space. X might have a different shape than depicted – all that matters is the proportion of $\overline{A}$ it occupies.

Now intuitively, the probability that the dart landed in B if it landed in A^+ is $\frac{1}{3}$, as one can see plainly in diagram 5: A^+ has a probability of $\frac{3}{4}$ and B of $\frac{1}{4}$. This could only happen if a third of the $\overline{A^+}$ points were most similar to a point in B. However we carefully constructed A^+ in such a way so that the only points where the most similar A^+-points lie in B are in fact just the points lying in B itself. Suppose the closest point to p in A^+ is in B. Due to the nature of similarity, this point is also the closest point to p lying in A. Thus the closest point to p lying in A lies in B, and since this is the defining condition for A^+, p must belong to A^+. Since p is in A^+, the point that's closest to p in A^+ is just p itself. Since by assumption this closest point lies in B, p lies in B as required.

It should be noted that for the purposes of modeling this situation I have identified each world in our model with the point on the dartboard where the dart landed. This effectively guarantees that the degree of similarity between two worlds is determined by where the dart lands in those two worlds: it may not be a function of the distance between their landing places, but it is a function of their landing places in some way or another. One could in principle work with more sophisticated models that take into account other features of the world – the weather in Shanghai, the temperature on Mars, and so on – but an intuitive explanation of how such modifications could affect our conclusion is hard to imagine.[53]

[53]There are purely formal models based on similarity ordering that can get the right probabilities (see the Stalnaker-Bernoulli models from van Fraassen [44]). However it is unclear to

Of course, there will be some who find the conclusion that the probabilities of conditionals go awry both unsurprising and not a problem exclusive to the similarity theorist: the various triviality results seem to indicate that these problems are unavoidable. I think this reaction is wrongheaded for several reasons. Firstly, I explicity challenge the idea that the triviality results warrant this conclusion. The triviality results are usually either directed at the similarity theorist and make assumptions that only they accept (such as in Stalnaker [36]), or are directed at an independently implausible version of the thesis.[54] In [XXX] I argue that one can in fact vindicate our pretheoretic intuitions about conditional probabilities using the kind of theory developed in section 1. Secondly, the problem of probabilities I have presented here affects only the similarity theorist – to get the problem I have appealed to principles that are specific to a similarity theorist, and are not part of the alternative picture we have suggested involving random selection. Lastly, the first objection was clearly on the mark – whatever the probability of that conditional is, it's not overwhelmingly likely that the dart landed on the edge if it landed in A. The revenge problem, even though not quite as striking, seems to be of a similar nature and arguably ought to recieve a similar diagnosis.

2.6 Against **CSO**: Iterated conditionals

A hallmark of a truth conditional theory is that it provides truth conditions for arbitrary embeddings of conditionals under various operations including embeddings under other conditionals. The predictions that a theory makes about embedded conditionals when compared with the data is often a useful source of evidence for or against that theory – for example, the behaviour of embedded conditionals plays a fairly substantial role in the arguments against the material account of indicatives (see Edgington [11] for an overview).[55] This is a notoriously difficult topic, and some have suggested that the lack of a satisfactory general theory of embedded conditionals indicates that a suppositional view, in which iterated conditionals aren't assigned meanings at all, would be more appropriate (see, for example, Edgington [11]).

I am inclined to find this type of attitude overly pessimistic. Conditionals embedded in other conditionals – especially when embedded in antecedent position – are often hard to make sense of, but they're not all impossible to interpret. An example due to Richmond Thomason (mentioned in van Fraassen [44]) demonstrates this quite nicely: if the plate broke if it fell on the floor, then it broke if it fell out of the window. It is perfectly good English, and we have no trouble evaluating what it means.

me how to relate this purely formal construction to the real-world application we've described.

[54] There are several triviality results that need discussion here. Since I address them in other work I shan't rehearse those arguments here.

[55] See, for example, Grice's proof of the existence of God: it's not the case that if God exists we may do whatever we like. Therefore God exists. The premise involves a conditional embedded under a negation and seems to be true, whereas the conclusion does not seem to follow. Adams' switch example, etc.

At any rate, the similarity theorist ought not to take this pessimistic attitude since their theory predicts that embedded conditionals are meaningful and even tells us what their truth conditions are. For the similarity and random world accounts of indicatives the assumption of meaningfulness is shared, so to get a sense of their comparative advantages it is extremely natural to compare their treatment of iterated conditionals with one another.

The cases I will focus on here concern the circumstances under which one can and can't infer something from a conditional. Typically, if you do not already know the truth of A or of B, you cannot infer A, or B, after learning the conditional 'if A then B'. To illustrate this point consider an instance of this type of fallacy, which will help prime us for the interesting case. Suppose that we have two coins, both of which have been flipped independently of one another. Now suppose Alice finds out something about the outcomes (perhaps that they're correlated), and informs Bob that if either coin landed heads, both did. Bob then reasons (fallaciously) as follows:

(1a) If either the first coin or the second coin landed heads, both coins landed heads.

(1b) So, both coins must have landed heads.

Again, we are to assume that Bob doesn't know anything about the outcomes of the coins before taking Alice's assertion into account. In particular he doesn't know whether the first or second coin landed heads. It seems clear that Bob is not in a position to infer that both coins landed heads once he learns the conditional – he is not in a position to infer the consequent (or the antecedent) of this conditional from the conditional alone.

Now consider what seems to be an analogous fallacy. Suppose that our original experiment has been modified slightly:

> There are two coins, the first of which we know has been flipped. We also know that if it landed heads the second coin was also flipped, and that otherwise the second coin was not flipped. Suppose now that this experiment has taken place: you know the parameters of the experiment, but do not know the outcome of any of the potential flips.

The only difference here is that the second coin might not have been flipped at all – it was only flipped if the first coin was flipped. Alice once again learns something about the outcome of the experiment and informs Bob that if either the first coin landed heads or the second coin landed heads if it was flipped, both coins landed heads. Bob then reasons as follows:

(2a) If either the first coin landed heads or the second coin landed heads if it was flipped, both coins landed heads.

(2b) So both coins landed heads!

It seems to me that just as in the previous example, Bob's conclusion is not warranted. There is simply no way for him to conclude that both coins landed heads since the antecedent is unknown to him: he doesn't know whether either coin landed heads if it was flipped.

In short, (2a) does not tell us that both coins landed heads, it only tells us that both coins landed heads *if* either the first coin landed heads (which it might not have) or the second coin landed heads if it was flipped (which also is not guaranteed: the second coin might land tails if it is flipped). Since Bob doesn't know the antecedent of this conditional – for all he knows the first coin didn't land heads, and the second landed tails if it was flipped – he can't conclude from (2a) that both coins landed heads. This is just the fallacy of infering the consequent of a conditional from the conditional itself.

This final judgment directly conflicts with the principle CSO, for given CSO, Bob is allowed to make exactly this inference. Just knowing that both coins landed heads if either the first coin landed heads or the second did if it was flipped is sufficient to conclude that both coins in fact *did* land heads! I put the formal argument in a footnote, but it effectively relies on the rule $A \to BC \vdash AB \to C$ which is a relatively straightforward consequence of CSO.[56] It is easy to see informally why the similarity theorist is committed to this entailment. Let us suppose that w is the closest world where either the first coin landed heads, or the second landed heads if it was flipped.[57] If the conditional (3) is true then, according to the similarity theory, the closest world w must be a world where both coins landed heads. Thus it follows that w is the closest world where the first coin landed heads, and thus the closest world in which the second coin was flipped. Here I am just appealing to a general fact about similarity: if the closest $A \vee B$ world is an A world, then that world is also simply the closest A world. So the closest world in which the second coin was flipped is a world in which it landed heads, namely w. So it's true at the actual world that the second coin landed heads if it was flipped, which by disjunction introduction tells us that the antecedent of (3) is true at the actual world. Since by assumption (3) is true at the actual world, if follows that both coins landed heads at the actual world by modus ponens.

Insofar as there is an appearance of invalidity here, one should try to find an explanation of that appearance. Of course, there are many avenues open to the similarity theorist in this regard, many of which emerge in the context of more general considerations about the nature of embeded conditionals. However

[56] Our conditional has the form $(A \vee (A \to B)) \to AB$, which by the consequence of CSO mentioned above entails $((A \vee (A \to B)) \wedge A) \to B$. The antecedent of this conditional is just equivalent to A by propositional logic, so making that substitution we can infer from this $A \to B$. Finally by disjunction introduction we have $A \vee (A \to B)$, and by modus ponens on our original conditional $(A \vee (A \to B)) \to AB$ we can infer AB. (In this proof I have formalized 'the second coin is flipped' and 'the first coin landed heads' using the same letter A, since we know they are equivalent given the setup of the experiment. This assumption can be put explicitly as a premise, or we can change the example slightly so that we substitute 'the second coin was flipped' with 'the first coin landed heads' or *vice versa*.)

[57] Very little changes in this argument if we replace w with a collection of closest worlds; this kind of argument works for people who deny CEM as well. Due to failures of the limit assumption, Lewis's semantics does not even reference a set of closest worlds so this argument doesn't directly apply to Lewis's semantics, however the proof theoretic argument in footnote 56 shows that even Lewis is committed to this conclusion.

the proponent of the random selection semantics has at least a *prima facie* advantage in that they can give a completely straightforward explanation of this appearance: the inference from 2a to 2b seems to be invalid because it *is* invalid.

This argument was stated as a problem for the similarity theorist. But it is important to stress that the problem does not rest on any potentially disputable hypotheses about the nature of similarity. It is a purely structural problem that applies to any theory that analyses conditionals in terms of an ordering over worlds of some kind, regardless of what those orderings represnt. Indeed by footnote 56 it applies to any theory that accepts the principle CSO, whether it be a theory based on similarity orderings or not.

The final example, of course, involved a conditional with a conditional embedded in its antecedent. The antecedent of this conditional is:

(2c) Either the first coin landed heads or the second coin landed heads if it was flipped

The hypothesis itself, I hope, is fairly straightforward. In the set up described, we can formulate a number of hypotheses about how the coins landed: that the first coin landed heads, that the second coin landed heads if it was flipped, and so on. It strikes me that (2c) is also a perfectly acceptable hypothesis: we can wonder about whether it is true, and we can also wonder about what else is true *if* it is true.

Sometimes people are tempted to read the embedded conditional as a conjunction, reading (2c) instead as: ‘either the first coin landed heads or the second coin was flipped and landed heads’. This mistake is perhaps encouraged by the fact that in this scenario both disjunctions seem to be equiprobable. However a little reflection reveals that they do not mean the same thing: one could change the set up so that I was pretty confident that the first coin is double-tailed and pretty confident the second coin is double-headed. Then I’d be pretty confident that the second coin landed heads if it was flipped, and thus confident in (1), yet also pretty confident that the first coin landed tails, and thus that the second coin wasn’t flipped, making me pretty confident that the disjunction ‘either the first coin landed heads or the second coin was flipped and landed heads’ is false.

Admittedly these kinds of subtleties must be born in mind when evaluating embedded conditionals; it takes a trained ear to discern the potential pitfalls. Nonetheless, the fact that one has to be careful with these kinds of examples is not itself a reason to avoid taking these kinds of judgments as evidence once one has been careful about them.

Before we move on let me note, as an aside, that the above verdicts follow from a much more general principle that says something like the following: if you are in a position to infer the antecedent of a conditional from the conditional itself (and the conditional is epistemically possible), then you already know the antecedent. Talk of ‘knowledge’ in this context isn’t particularly precise, but note that we can define an natural epistemic modality from the conditional by setting $\Box A$ equal to the defined connective $\neg A \to A$. In a context where the accessibility relation E is salient, $\Box A$ so defined just expressed the modal operator generated by the accessibility relation E. Using this operator we can formulate the principle in question in the object language:

EA $(\Diamond(A \to B) \wedge \Box((A \to B) \supset A)) \supset \Box A$.

This principle, at least in its informal incarnation, has a degree of intuitive appeal motivated by examples like the above ones. More interestingly, it can also be justified by appeal to Stalnaker's thesis. This can be seen in two steps: (i) it is a completely general fact that if A is probabilistically independent of B, and the probability of A is non-zero, then you can be certain in the material conditional $A \supset B$ only if you are already certain in B (ii) Stalnaker's thesis entails that every conditional is probabilistically independent of its antecedent,[58] thus by (i) if a conditional has non-zero probability then you can be certain in the material conditional $(A \to B) \supset A$ only you are already certain in A. (For similar reasons, the principle EA has the virtue of always having probability 1 for a conditional that satisfies Stalnaker's thesis[59].) The combination of EA with CSO is disastrous, however: it collapses the conditional into the material conditional.

It should be emphasised, however, that nothing in the preceding discussion depended on the truth of this more general principle – I appealed only to direct intuitions concerning what can be inferred from conditionals.

3 Conclusion

The difference between the random selection semantics and the ordering semantics is a subtle one and turns on principles, such as CSO, whose validity is hard to evaluate directly. While we have seen a pattern of judgments emerging that best fits the hypothesis that CSO is invalid, these considerations are tentative and I think it would be wrongheaded to rest an entire theory on a few judgments of this sort. Ultimately such matters are to be settled by systematic theoretical considerations that bring into the deliberation broader bodies of theory.

The theory of random selection based on Proportionality is a simple general theory exhibiting exactly these sorts of theoretical virtues. It provides a systematic explanation of our judgments about the probabilities of indicative conditionals, whilst also correctly predicting the invalidity of AS and other monotonicity principles, and the validity of the indicative versions of CEM and LIM. By contrast the flatfooted version of the similarity theory, whilst providing a concrete model of the failure of monotonicity, cannot straightforwardly account for our judgments about probability and does not validate CEM or LIM. Perhaps this would not matter if one were looking only at the case of subjunctives, as Lewis was, however these considerations to appear to render the flatfooted version of the theory unsuitable for indicatives. On the other hand,

[58] $Pr(A \to B \mid A) = Pr(A \wedge (A \to B) \mid A) = Pr(AB \mid A) = Pr(B \mid A) = Pr(A \to B)$. The second identity is obtained by applying the law $((A \to B) \wedge A) \equiv (A \wedge B)$.

[59] Or, at least, a slight strengthening of it in which the probability of $A \to B$ must be 1 when the probability of A is 0.

the amendments one has to make to the flatfooted theory to ensure that it conforms to our judgments about indicatives, particularly with respect to principles like CEM and LIM, undermines many of those theoretical virtues. One has to give up the simple similarity test for evaluating indicatives for truth and one must appeal instead to 'ironed out' orderings – heavily altered orderings that bear little relation to the underlying ordering of worlds by similarity to actuality.[60]

The theory we have developed based on the theory of random selection completely eschews the notion of similarity in favour of a probabilistic theory of conditionals very much in line with the program of Stalnaker 1970 [38]. What is interesting about our present theory is that while it was explicitly developed with probabilistic considerations in mind, it also appears to have the resources to respond to many of the problems with the similarity analysis alluded to above.

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

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

[5] Wylie Breckenridge and Ofra Magidor. Arbitrary reference. *Philosophical Studies*, 158(3):377–400, 2012.

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

[7] Cian Dorr and John Hawthorne. Naturalness. 2013.

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

[9] Kenny Easwaran. Regularity and hyperreal credences. *Philosophical Review*, 123(1):1–41, 2014.

[10] D. Edgington. Indicative conditionals. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Winter 2014 edition, 2014.

[60] The amended theory thus seems to be much less predictive than the flatfooted version, for one has chosen the ordering to fit the data rather than using an antecedently given ordering to make predictions about the data.

[11] Dorothy Edgington. Do conditionals have truth conditions? *Crítica: Revista Hispanoamericana de Filosofía*, 18(52):3–39, 1986.

[12] Kit Fine. Critical notice of lewis, counterfactuals. *Mind*, 84(335):451–458, 1975.

[13] Kit Fine. A difficulty for the possible worlds analysis of counterfactuals. *Synthese*, 189(1):29–57, 2012.

[14] 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.

[15] Anthony S Gillies. Counterfactual scorekeeping. *Linguistics and Philosophy*, 30(3):329–360, 2007.

[16] J. Goodman. Consequences of conditional excluded middle. *Unpublished*.

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

[18] 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.

[19] William L Harper. Ramsey test conditionals and iterated belief change (a response to stalnaker). In *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, pages 117–135. Springer, 1976.

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

[21] Hans G. Herzberger. Counterfactuals and consistency. *Journal of Philosophy*, 76(2):83–88, 1979.

[22] Boris Kment. *Modality and explanatory reasoning.* OUP Oxford, 2014.

[23] Angelika Kratzer. Conditionals. In *Chicago Linguistics Society*, volume 22, pages 1–15, 1986.

[24] Saul A Kripke. A completeness theorem in modal logic. *The journal of symbolic logic*, 24(01):1–14, 1959.

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

[26] D. Lewis. A subjectivist's guide to objective chance. *Studies in inductive logic and probability*, 2:263–293, 1980.

[27] David Lewis. Counterfactuals and comparative possibility. In *Ifs*, pages 57–85. Springer, 1981.

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

[29] Sarah Moss. On the pragmatics of counterfactuals. *Noûs*, 46(3):561–586, 2012.

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

[31] M. Schulz. Counterfactuals and arbitrariness. *Mind*, 123(492):1021–1055, 2014.

[32] Brian Skyrms. Causal necessity: a pragmatic investigation of the necessity of laws. 1980.

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

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

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

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

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

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

[39] R.C. Stalnaker. *Inquiry*. 1984.

[40] Robert Stalnaker. *Context and content: Essays on intentionality in speech and thought*. 1999.

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

[42] Pavel Tichỳ. A counterexample to the stalnaker-lewis analysis of counterfactuals. *Philosophical Studies*, 29(4):271–273, 1976.

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

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

[45] Robert Van Rooy. Gibbard's problem: The context dependence of conditional statements. *Proceedings Dutch-German workshop on non-monotonic reasoning*, 1999.

[46] Kai Von Fintel. Counterfactuals in a dynamic context. *Current Studies in Linguistics Series*, 36:123–152, 2001.

[47] J Robert G Williams. Chances, counterfactuals, and similarity. *Philosophy and Phenomenological Research*, 77(2):385–420, 2008.

[48] J Robert G Williams. Defending conditional excluded middle. *Noûs*, 44(4):650–668, 2010.