Conditionals: Truth, Success and Safety

D H MELLOR & RICHARD BRADLEY

Abstract

The paper develops a theory of conditionals based on how our conditional beliefs affect what we do as a means to an end, and how their truth makes those actions succeed in achieving their ends. It does this by taking the conditionals we act on to be true iff the inferential dispositions they express are contingently truth-preserving ('safe' for short), i.e. iff in the circumstances their consequents will or would be true if their antecedents are or would be true. It then extends the theory to other kinds of conditionals: third-person, indicative and subjunctive, past-referring, non-contingent, and complex. Finally, it gives a semantics on which identifying the truth values of conditionals with the 'safety values' of the inferential dispositions they express makes them behave like truth values in truth functions with other conditional and unconditional statements.

1. Introduction and summary

Whether conditionals have truth values and, if so, what they are, are still matters of controversy. The principal answers, when 'P' and 'Q' are unconditional, true or false, and logically independent, are that 'If P then Q'

A: has no truth value (Adams 1975, Edgington 1986, Gibbard 1981– for indicative conditionals);

B: is true when 'P' is true, if and only if 'Q' is true; and when 'P' is false, either

B1: has either a third or no truth value (Belnap 1970, McDermott 1996, Milne 1997, Bradley 2002, Edgington 2005); or

B2: is true (Jackson 1990, Lewis 1976 – for indicative conditionals); or

B3: is true if 'Q' is true in all the accessible[1] P-worlds (i.e. worlds where 'P' is true) that rank highest in some strict ordering of P-worlds (e.g. by how closely they resemble our ¬P-world), and false if it isn't (Lewis 1973, Stalnaker 1984, Kratzer 2012).

In this paper we derive a different answer from the two-part theory that

(I) To *believe* 'If P then Q' is to have a disposition, D(P>Q), to infer 'Q' from 'P',[2] and

---

1 'Accessible' in the sense of Lewis (1973, §1.2): 'Necessity of a certain sort is truth at all possible worlds that satisfy a certain restriction. We call these worlds accessible, meaning thereby simply that they satisfy the restriction.' Our only restriction is that, pace dialetheism (Priest et al. 2018), in no accessible world is any proposition both true and false.

2 This combines the 'suppositional' view, that to believe 'If P then Q' is to be disposed to believe 'F' on the supposition that 'E' is true (Adams 1965, Gärdenfors 1986, Edgington 2009), with the view that this disposition is inferential (Stalnaker 1984, Mellor 1993): i.e. that it is the state of mind which makes believing 'P' cause us to believe 'Q'.

(II) 'If P then Q' is *true* iff in the circumstances D(P>Q) is truth-*preserving* or, for short, *safe*: i.e. iff D(P>Q) wouldn't in the circumstances make a true belief in 'P' cause a false belief in 'Q'.[3]

Our core argument for (I) and (II) is that they, unlike A–B3, can account for the role of conditionals in decision making. We start in §2 by saying how we take conditional beliefs to affect what we do as a means to an end. Thus suppose for example I'm deciding whether to pay £1 (P) to get a newspaper (Q). Then whether I make 'P' true by paying, or false by not paying, I do so because, *inter alia*, I believe both 'If P then Q' and 'If ¬P then ¬Q': i.e., given (I), because I'm disposed both to believe 'Q' if I come to believe 'P' (by making it true), and to believe '¬Q' if I come to believe '¬P' (by making 'P' false).

In §3 we infer from (II) that for either of these actions (making 'P' true or making 'P' false) to *succeed* as a sufficient and necessary means to its end (making 'Q' true or making 'Q' false), 'If P then Q' and 'If ¬P then ¬Q' must both be *true*: a condition that requires both of them to be true whether their antecedents are true or false. This requirement is easily met by our thesis (I), that conditionals express inferential dispositions, since dispositions are generally independent of whether they are being manifested: salt, for example, is as water-soluble – disposed to dissolve in water – out of water as it is in it. However, as we show in §4, the requirement can't be met by any of the theories A–B3 listed above.

In §5, we show how to apply our theory of conditional truth to conditionals of all kinds, not just those we act on; and we conclude in §6 by giving a semantics on which the truth values our theory gives conditionals do indeed behave like truth values in all truth functions with other conditional and unconditional constituents.

2. Acting on Conditionals

For present purposes we need only consider how our actions are affected by conditionals like 'If P then Q' of which we are practically *certain*, i.e. where we are disposed to *fully* believe 'Q' if we believe 'P'. The decision theories of Ramsey (1926), Jeffrey (1965), *et al*. of course apply also to acting under *uncertainty*, where our credence (probabilistic degree of belief) in 'Q' if we believe 'P' is sufficiently less than 1 to affect how, in the circumstances, we will or would act on it. Still, even if acting under practical certainty is only a special case, it is also a very common one: when deciding whether to pay for newspapers, or for any other goods or services, I'm usually quite sure I will get them if I pay for them and won't if I don't.

Confining ourselves to action under certainty will both simplify what follows and enable us to evade disputes about the role and kinds of probability involved in decision making under uncertainty. Another dispute we can also evade is whether decision theories should be read descriptively, as saying

---

3 'Safe' here doesn't mean 'valid', i.e. *necessarily* safe: when 'P' and 'Q' are logically independent, D(P>Q)'s safety, and hence 'If P then Q's truth, are contingent, e.g., when 'P' is 'I go out' and 'Q' is 'I get wet', on the weather.

how our conditional beliefs *will* affect our actions (Blackburn 1998 ch. 6; Mellor 2005), or prescriptively, as saying how they *should* affect them (Joyce 1999), perhaps because if they didn’t we’d be irrational, which it’s assumed we shouldn’t be. As for present purposes either reading will do, prescriptivists should in what follows read ‘will’ as ‘should’ where appropriate.

Suppose then, to take a more adaptable example, that I have to decide, when booking a long-haul flight, whether to pay extra (E) to fly 1st class (F). Then whether I will make ‘E’ true depends on two factors:

(a) what I want to happen, and

(b) how I believe the class I will fly in depends on what I do.

The four relevant scenarios in (a) are:

E&F, I pay extra and fly 1st;

E&¬F, I pay extra but don’t fly 1st;

¬E&¬F, I don’t pay extra and don’t fly 1st; and

¬E&F, I don’t pay extra but fly 1st anyway.

And as I much prefer the last of these scenarios, ¬E&F, only if something rules it out will I consider paying extra to fly 1st, i.e. making ‘E’ true in order to make ‘F’ true.

What makes me rule out ¬E&F is the other factor, (b), that my decision depends on: namely, my belief that if I do pay extra I *will* fly 1st and if I don’t I *won’t*, i.e. that if E then F and if ¬E then ¬F. These beliefs are what, by reducing my foreseeable scenarios to E&F and ¬E&¬F, and thereby ruling out ¬E&F, will cause me to

make ‘E’ true if I think it *is* worth paying the extra in order to fly 1st, or

make ‘E’ false if I think it’s *not* worth doing so.

In other words, my believing ‘If E then F’ and ‘If ¬E then ¬F’ will cause me to make ‘E’ true if I prefer E&F to ¬E&¬F, and to make ‘E’ false if I prefer ¬E&¬F to E&F.[4] Thus, in tabular form, if X is whatever else it takes to book my flight:

Conjunctively Sufficient Causes of Action				Action
X	I believe ‘If E then F’	I believe ‘If ¬E then ¬F’	I prefer E&F to ¬E&¬F	I make ‘E’ true
			I prefer ¬E&¬F to E&F	I make ‘E’ false

Table 1

4 This causation will only be *sufficient*, not *necessary*, since even if I am indifferent between E&F and ¬E&¬F, I will still have to decide whether to pay extra to fly 1st: in which case I will either make ‘E’ true without positively preferring E&F to ¬E&¬F, or make ‘E’ false without positively preferring ¬E&¬F to E&F.

Table 1 shows that, whichever action I take, one cause of it will be that I believe both ‘If E then F’ and ‘If ¬E then ¬F’, and therefore that I will believe both of them whether I make ‘E’ true or false.

3. Success and Truth

The alternative actions in Table 1 will *succeed*, i.e. cause their respective ends, if I will fly 1st if I pay extra and won’t if I don’t. If this is so, then the dispositions D(E>F) and D(¬E>¬F) that (I) says ‘If E then F’ and ‘If ¬E then ¬F’ express will be what we shall call ‘*safe*’: i.e., they won’t make a true belief in ‘E’ cause a false belief in ‘F’ or a true belief in ‘¬E’ cause a false belief in ‘¬F’.[5] And that, our thesis (II) says, is what makes those conditionals *true*. Thus, in tabular form:

Conjunctively Sufficient Causes of Success		Action	Effect
‘If E then F’ is true	‘If ¬E then ¬F’ is true	I make ‘E’ true	E&F
		I make ‘E’ false	¬E&¬F

Table 2

However, Table 2’s assumption that ‘If E then F’ and ‘If ¬E then ¬F’ can be true together is contentious. Some theories (e.g. B1-3 above) agree that each will be true if, when its antecedent is true, i.e. when it’s *actual*, its consequent is also true. What’s contentious is whether either of them can be true when its antecedent is *false*, i.e. when it’s *counter*-actual, as one of them must be. One objection to this is based on the fact that if, for example, I make ‘E’ false, nothing in our actual ¬E-world will determine *which* non-actual E-world my making ‘E’ true would take me to; the other is based on the non-zero *chance* that a non-actual E-world might then have of *not* being an F-world. We respond to these objections in turn.

3.1 *Undetermined possible worlds*

We agree that, when ‘E’ is false, nothing in our ¬E-world determines which of all the E-worlds accessible to ours my making ‘E’ true would take me to. Still, however many accessible E-worlds my making ‘E’ true *could* take me to, it couldn’t take me to *more* than one: since every proposition whose differing truth values differentiate those E-worlds would then be both true and false, which it can’t be.[6] And in whatever single E-world my making ‘E’ true *would* take me to, either I will fly 1st in it or I won’t: so either ‘F’ will be true in that world or ‘¬F’ will be.

5 We take for granted, *pace* Tooley (1997) and other ‘growing block’ theorists, that unconditional future-referring sentences like ‘I’ll fly 1st’ have truth values.

6 See footnote 1.

But if ‘F’ *is* true in that world, i.e. if ‘F’ *would* be true if ‘E’ *was* true, then the counter-actual ‘If E then F’ will be true; just as, if ‘E’ is true, the actual ‘If E then F’ will be true if ‘F’ is true. So ‘If E then F’ can be true both when it’s actual and when it’s counter-actual, as Table 2 requires; and so, for the same reason, can ‘If ¬E then ¬F’.

3.2 *Chance and truth*

Suppose now that if I did pay extra, I would still have a non-zero chance *ch*(¬F) of *not* flying 1st, perhaps because there’d be a non-zero chance of engine failure grounding my plane. Hawthorne (2005 p. 396) and others, using other examples, say that believing this, i.e. believing

‘If E then $ch(\neg F)=p$’ for some p between 0 and 1, and hence, since $ch(\neg F) = 1-ch(F)$,

‘If E then $ch(F)= 1-p$’,

will make us *dis*believe both the counter-actuals

‘If E then F’ and ‘If E then ¬F’.

Does this, if true, make the counter-actual ‘If E then $ch(\neg F)=p$’ incompatible with ‘If E then F’ and ‘If E then ¬F’? That is, does it entail that, when I don’t pay extra, it’s false both that I would, and that I wouldn’t, have flown 1st if I had paid extra?

Fortunately not: ‘If E then $ch(\neg F)=p$’ *isn’t* incompatible with ‘If E then F’ and ‘If E then ¬F’ when ‘E’ is false. This follows from the ‘laws of large numbers’ (Feller 1957 p. 141) that link chances to the frequencies they’re invoked to explain, and from which their values may, within limits, be inferred. These links range from a radium atom’s chances of decaying in given times and the proportions of them that do so, to, in our case, the chances of flights being grounded (G) and how often they *are* grounded.

In our example the relevant link is this: ‘$ch(G)=p$’ implies that some feature S of my flight (e.g. the age of the plane’s engines) is such that if, of $n(S)$ flights that are independently S, $f_{n(S)}(G)$ is the fraction that are grounded, then

as $n(S)\rightarrow\infty$, $f_{n(S)}(G)$’s chance of differing from p by any given amount ε, however small, → 0.

But this can only be true if, for any finite $n(S)$, $f_{n(S)}(G)$ *has* a value whose chance of differing from p by more than ε will tend to zero as $n(S)$ increases without limit. And since in this case, as in most others, the total number $N(S)$ of *actual* S-flights is finite, $n(S)$ can only increase without limit by eventually including merely *possible* S-flights. The value of $f_{n(S)}(G)$ that satisfies the law of large numbers will then be a fraction of actual and possible S-flights, a fraction that can only exist if *each* of those S-flights is either G or $\neg G$, i.e. if, for all x, either

‘If x is an S-flight then x is G’ is true or ‘If x is an S-flight then x is $\neg G$’ is true,

whether those conditionals are actual or counter-actual.

Similarly for ‘If E then F’ and ‘If E then ¬F’: the truth of ‘If E then *ch*(¬F)=*p*’, far from making these conditionals false when they’re counter-actual, entails that, as we argue in §3.1, one of them will be true even then. So a non-zero chance of not flying 1st if I pay extra does *not* stop ‘If E then F’ being true whether ‘E’ is true or false, as Table 2 requires.

How then does ‘If E then *ch*(¬F)=*p*’ conflict with ‘If E then F’ and ‘If E then ¬F’, as Hawthorne *et al.*, say it does? We think it does so by entailing that if ‘E’ was true, ‘F’ *might* be false, so that saying ‘I *might not* fly 1st’ suggests that I don’t *know* that I *would* fly 1st (and similarly for ‘I might fly 1st’ and ‘I wouldn’t fly 1st’). That’s why saying ‘If E then *ch*(¬F)=*p*’ has the ‘conversational implicature’ (Grice 1975) that ‘If E then F’ and ‘If E then ¬F’ are both false, despite entailing no such thing.

4. Theories A–B3

We showed in §2 how conditionals we fully believe affect what we do as a means to an end, and in §3 that what makes those actions succeed in achieving their ends is that those conditionals will be true whether they are actual or counter-actual. We now show why none of the theories A–B3 in §1 can explain these results.

A’s giving the conditionals we believe *no* truth values stops it explaining what makes the means-to-end actions they affect succeed in achieving their ends.

B1’s giving all *counter*-actual conditionals no (or a third) truth value stops ‘If E then F’ and ‘If ¬E then ¬F’ being true together, which, as Table 2 shows, is what will make ‘E’ truth cause ‘F’ to be true or ‘E’s falsehood cause ‘F’ to be false.

The trouble with B2’s truth-functional ‘material’ conditionals isn’t that they’re *not* true when they’re counter-actual but that they’re *all* true then. This makes *both* ‘If ¬E then F’ *and* ‘If ¬E then ¬F’ true when ‘E’ is true, which for the reason given in §3.1 they can’t be: for in no accessible ¬E-world that my making ‘E’ false could take me to can ‘F’ and ‘¬F’ both be true. Nor, given (I), could I *believe* ‘If ¬E then F’ and ‘If ¬E then ¬F’ simultaneously, since I can't be disposed to believe ‘F’ *and* ‘¬F’ if I believe ‘¬E’. In short, B2 can explain neither why believing ‘If ¬E then ¬F’ but *not* ‘If ¬E then F’ will cause me to make ‘E’ true in order to make ‘F’ true, nor why I need to make ‘E’ true to make ‘F’ true because ‘If ¬E then ¬F’ is true and ‘If ¬E then F’ isn’t.

B3 theories, which say that when ‘E’ is false, ‘If E then F’ will be true in our ¬E-world iff

> ‘F’ is true in *all* the non-actual E-worlds that rank highest in some strict ordering of them,

can also let ‘If E then F’ be true whether ‘E’ is true or false, as Table 2 requires, and similarly for ‘If ¬E then¬F’. Our objection to these theories is that what *makes* ‘If E then F’ true when ‘E’ is false isn’t any *overall* ranking of non-actual E-worlds, e.g. their overall similarity to our ¬E world (Lewis 1973, ch. 2.3).

To show this, we recall first that most of the conditionals we act on are contingent.[7] Thus, in our flying example, the truth of 'If E then F' and 'If ¬E then ¬F' is contingent on the laws of nature, and global facts like the atmosphere's density, that enable aircraft to fly, the conjunction of which we'll call 'Y'. It's also contingent on many variable local facts – the airline's pricing policy, the availability of 1st class seats, the plane's taking off, etc. – the conjunction of which we'll call 'Z'. And what determines whether a non-actual E-world is also an F-world isn't how similar to our ¬E-world it is *overall*, but whether 'Y&Z' is true in that E-world, however much it differs from our ¬E-world in other ways.

Compare for example an E-world w_1 where I pay extra and fly 1st (since 'Y&Z' is true), but on a plane which then crashes and kills us all, with an E-world w_2 where I pay extra but don't fly at all, because the flight is cancelled (making 'Z' and hence 'Y&Z' false). What makes 'F' true in E-world w_1 and false in E-world w_2 is that 'Y&Z' is true in w_1 and false in w_2: the fact that w_1 resembles our world far less overall than w_2 does is irrelevant. In short, what makes 'If E then F' true when 'E' is false is nothing more general than Y&Z's independence of E.

5. Kinds of Conditionals

Having shown how our theory, unlike theories A–B3, makes sense of how the conditionals we act on affect our actions, and their truth makes those actions succeed, we must now show how the theory applies to other kinds of conditionals.

5.1 *Third-person conditionals*

The conditionals we act on are future-referring first-person ones like 'If I pay extra I'll fly 1st' or, to vary the example, a conditional that our fit friend Anna, wondering whether to run a marathon, believes:

'If I run I'll finish'.

Still, what's true of this conditional is also true of its impersonal counterpart,

C1: 'If Anna runs, she'll finish',

which we believe whether or not we believe Anna will run, and which, since Anna's fitness makes D(Anna runs > Anna finishes) safe whether she runs or not, will be *true* whether she runs or not.

5.2 *Indicative and subjunctive conditionals*

Our belief in future-referring conditionals isn't always independent of whether we believe their antecedents, and their truth isn't always independent of their antecedents' truth. Suppose for example that, since we believe Anna wouldn't run if she was too unfit to finish, we believe

C2: 'If Anna *does* run, she *will* finish'

7 See footnote 3.

even though, because we also believe Anna is too ill to finish, we *dis*believe

C3: 'If Anna *were* to run she *would* finish'.

This distinction, between an indicative reading, C2, and a subjunctive reading, C3, of

C1: 'If Anna runs she'll finish'

is admittedly contentious, being accepted by some philosophers (Gibbard and Harper 1978, Joyce 1999, Bradley 2017) but rejected by others (Dudman 1988, Bennett 1988). Here, while we needn't take sides, we do need to account for the distinction if there is one. To that end, in order to distinguish C2 from the equally indicative C1, we shall call C2 'evidential', and C3 and C1 'causal', with C3 differing from C1 by being counter-actual, i.e. by implying, unlike C1, that Anna doesn't run.

The problem simultaneous belief in C2 and disbelief in C3 poses for our theory is that C2 and C3 both express the same disposition, D(Anna runs > Anna finishes), which in no accessible world can we both have (as our belief in C2 implies) and lack (as our disbelief in C3 implies) at the same time.[8] The solution is that, knowing Anna, we will only infer 'Anna finishes' from 'Anna runs' if we believe 'Anna runs', an inference which our believing Ann's too ill to run would otherwise stop us drawing.

In other words, our beliefs about Anna stop our D(Anna runs > Anna finishes) being causally independent of our belief in 'Anna runs', as it would be if we believed C1 because we believed Anna was fit. That's what makes C2 both actual and evidential: only evidence that 'Anna runs' is actually true will dispose us to infer that Anna finishes. Without this evidence we won't be disposed to draw that inference, which is why we disbelieve the causal and counter-actual C3. And as for belief, so for truth: what makes the actual and evidential C2 true is that Anna won't run unless she's fit enough to finish; what makes the causal and counter-actual C3 false is that she's not that fit.

5.3 *Past-referring conditionals*

Suppose first that Anna's fitness does make the future-referring C1, 'If Anna runs she'll finish', true whether its antecedent is true or false. Then both of C1's past-referring counterparts, the evidential

C2P: 'If Anna did run, she did finish',

and the causal

C3P: 'If Anna had run, she would have finished',

will also be true.

But now suppose, as in §5.2, that Anna's illness makes C3 false, while her common sense makes C2 true. In that case the past-referring counterparts of these future-referring conditionals will again share those conditionals' truth values: the evidential C2P will be true, and the causal C3P will be false.

8 See footnote 1.

Similarly for the truth values of the well-known past-referring conditionals about Lee Harvey Oswald's assassination of the US President John F. Kennedy. We will only believe the causal and counter-actual

'If Oswald hadn't killed Kennedy someone else would have'

if we believe Oswald had backup, since we know that only then would the disposition

D(Oswald didn't kill Kennedy > someone else did)

be safe. Whereas, since we know that Kennedy *was* killed, we do believe the evidential and actual

'If Oswald didn't kill Kennedy someone else did'

whether or not we believe he had backup.[9]

5.4 *'Centred' conditionals*

Another debatable consequence of our theory, as of B1-3, is the 'centering' principle (Lewis 1973 p. 14) that, for all 'P' and 'Q', 'If P then Q' will be true if 'P' and 'Q' are true. This is uncontentious when P and Q are causally linked, as my paying extra and flying 1st are by the facts Y&Z cited in §4, and Anna's running and finishing are by her fitness. The principle is more contentious when they aren't so linked, as in the conditional, 'If London is large, water is wet', which sounds very odd. Yet our theory makes it true because, since London *is* large and water *is* wet, the inferential disposition it expresses is in fact safe.

What makes this conditional sound odd is not only that its consequent is too well known to need inferring from its antecedent, but that we wouldn't assert it anyway: because doing so would tacitly imply something false, namely that we believe London's largeness makes water wet, which we don't. But this no more shows that 'If London is large, water is wet' is untrue than, as we saw in §3.2, the conversational impropriety of conjoining 'might not' and 'would' counter-actuals shows that the former's truth falsifies the latter.

5.5 *Non-contingent conditionals*

(i) A contingent 'P' that entails a contingent 'Q', thereby making it necessarily safe to infer 'Q' from 'P', and unsafe to infer '¬Q', makes 'If P then Q' necessarily true and 'If P then ¬Q' necessarily false.

(ii) If 'P' and 'Q' are both necessarily true, then so is 'If P then Q', because the inference from 'P' to 'Q' is again necessarily safe. Some of these conditionals admittedly sound as odd as §5.4's 'If London is

---

9 In this case what makes the evidential conditional true, and makes us believe it, is not the assumed truth of its antecedent, which is what makes C2P true, but the known truth of 'Someone killed Kennedy' and hence of 'Oswald or someone else killed Kennedy'. Our belief in that is what makes us believe 'If Oswald didn't kill Kennedy someone else did', by disposing us to infer its consequent from its antecedent: a disposition that's only made safe by the fact that someone did kill Kennedy. Had Kennedy not been killed, that inference might not have been safe, which is what limits its known safety, and the known truth of the conditional that expresses it, to the actual world, in which we know he was killed.

large, water is wet', and for the same reason: asserting 'If 2+2=4, 3 is prime', for example, suggests, falsely, that we think the truth of '3 is prime' depends on that of '2+2=4'. But as in the contingent case, so in this necessary one: the oddness of asserting it is no reason to deny its truth.

(iii) If 'P' is necessarily true and 'Q' is necessarily false, as in 'If 2+2=4, 3 is not prime', the inference from 'P' to 'Q' will be necessarily unsafe, and 'If P then Q' therefore necessarily false.

(iv) If 'P' is necessarily false, the impossibility of its truth makes inferring 'Q' from 'P' trivially safe, since it can't make a true belief in 'P' cause a false belief in 'Q'. The fact that this, on our theory, makes 'If P then Q' necessarily true for all 'Q', may however seem to vitiate our theory, e.g. of *reductio* proofs of the form 'If P then ¬P; therefore ¬P', like

'If there is a greatest prime number, p_N, then

$p_1p_2\ldots p_N + 1$ either is, or is divisible by, a prime number $> p_N$;

therefore there is no greatest prime number'.

The problem is that proofs like this will only be valid if the inference from 'P' to '¬P' is necessarily safe, and we can only believe them by believing that 'If P then ¬P' is necessarily true. Yet how could believing 'P' cause us to believe '¬P, when believing 'P' entails *dis*believing '¬P', which in no accessible world can anything cause us to believe and not believe simultaneously?[10] The answer is that, since believing 'P' couldn't cause us to believe '¬P', all that believing 'If P then ¬P' can do is make us disbelieve 'P', which is after all what a *reductio* proof is meant to make us do.

5.6 *Complex conditionals*

On our theory, a conditional 'If P then Q' with unconditional 'P' and 'Q' is made true, if it is, by the safety of the inference from 'P' to 'Q'. How can this apply to conditionals one or both of whose constituents are themselves conditionals: e.g., to revert to our flying example,

'If I fly Virgin, I'll fly 1st if I pay extra' or, for short, 'If V, then (if E then F)', or

'If I'll fly 1st if I pay extra, I'll fly Virgin,' or, for short, 'If (if E then F), then V'?

Our answer relies on the realism about dispositions (Armstrong 1968, Mellor 2000) tacitly implied in Table 1 by the causal roles of our believing 'If E then F' and 'If ¬E then ¬F', i.e. by our having the inferential dispositions D(E>F) and D(¬E>¬F). This realism, which we here take for granted, and takes our inferential dispositions to be real states of mind, with real causes and effects, lets us take

'If V, then (If E then F)'

to express a disposition

D(V>D(E>F))

10 See footnote 1.

that makes believing 'V' cause me to believe 'If E then F'. But then, if D(E>F) will be safe if 'V' is true, D(V>D(E>F)) will also be safe, since it won't then make true beliefs in 'V' and 'E' cause a false belief in 'F'. And if and only if that is so will the conditional 'If V, then (if E then F)', which expresses D(V>D(E>F)), be true.

Similarly, *mutatis mutandis*, for 'If (If E then F), then V', i.e. 'If I'll fly 1st if I pay extra, I'll fly Virgin'. The disposition this conditional expresses, D(D(E>F)>V), will make D(E>F) cause me to believe 'V'. That will then make D(D(E>F)>V) safe, and 'If (if E then F), then V' true, if and only 'V' will be true if 'If E then F' is true. And so on, to any graspable level of complexity: our theory will make any complex conditional we can understand as true or false as the simple conditionals that are its ultimate conditional constituents.

6. Semantics

We have shown in §5 how the truth values of all kinds of conditionals correspond credibly with the safety values of the inferential dispositions they express. But this isn't enough to show that for conditionals to be true *is* for the dispositions they express to be safe. To show that, we need to show that the truth values thus determined *behave* like truth values in truth functions with other conditional and unconditional constituents.

To show this, we note first that our theory makes 'If P then Q' true just in case, whenever 'P' is or would be true, 'Q' will or would also be true. In other words, this conditional's truth value depends on the state of the world either as it is, if 'P' is true or, if 'P' is false, as it would be if 'P' were true. To accommodate this we adopt Bradley's (2012, §4) multi-dimensional possible-world semantics, on which the truth value of 'If P then Q' varies not with single worlds but with *pairs* of them.

Thus suppose w is the actual world, or another accessible world, and w_E is a P-world accessible to w. Then, on this semantics, 'If P then Q' will be true at a pair of worlds (w, w_E) just in case 'Q' is true at w_E. This covers 'If P then Q' both when it's *actual*, i.e. when 'P' is actually true, and when it's *counter*-actual, when 'P' is actually false. For when 'P' is *true* in w, and $w_E = w$, 'If P then Q' will be actual in w, and true there just in case 'Q' is true there. Whereas, when 'P' is *false* in w, so that w is *not* a P-world w_E, 'If P then Q' will be counter-actual in w, and true at just the world pairs (w, w_E) where 'Q' is true in w_E.

The relevant consequences of this semantics are as follows. Firstly, it makes the *conjunction* of two conditionals, 'If P then Q' and 'If R then S', true if and only if both are true. For on our theory, 'If P then Q' and 'If R then S' will be true just in case both inferences, from 'P' to 'Q' and from 'R' to 'S', are safe. So our semantics will, as required, make the conjunction of any two conditionals true if and only if both conditionals are true.

Secondly, the semantics will, as it must, make the *disjunction* of two conditionals true if and only if at least one of them is true. For it will make that so just in case at least one of the inferences those conditionals express is safe.

Finally, we know that 'If P then Q' will be true if and only if 'Q' is or would be *true* if 'P' is or was true, and that 'If P then ¬Q' will be true if and only if 'Q' is or would be *false* if 'P' is or was true. So the set of world pairs (w, w_E) where 'P' is true at w and 'Q' is *false* at w_E form the complement, within the set of all world pairs, of those where 'P' is true at w and 'Q' is *true* at w_E. This makes 'If P then ¬Q' equivalent to the negation of 'If P then Q', thereby satisfying the law of conditional excluded middle, by making 'If P then Q or if P then ¬Q' a logical truth.

The fact that this semantics for our theory meets these three desiderata shows that all Boolean compounds of conditionals will, as they should, have truth conditions that are Boolean functions of the truth conditions of their constituents. And that, we think, completes an overwhelming case for holding that the truth values of conditionals are determined by the safety values of the inferential dispositions they express.

References

Adams, E. (1975) *The Logic of Conditionals*. Dordrecht: Reidel.

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*. London, Routledge.

Belnap, N. (1970) Conditional Assertion and Restricted Quantification. *Noûs* 4: 1–13.

Bennett, J. (1988) Farewell to the Phlogiston Theory of Conditionals. *Mind* 97: 509–527.

Bradley, R. (2002) Indicative Conditionals. *Erkenntnis* 56: 345–378.

— (2012) Multidimensional Possible-World Semantics for Conditionals. *Philosophical Review* 121: 539–571.

— (2017) *Decision Theory with a Human Face*. Cambridge, Cambridge University Press.

Dudman, V. H. (1988) Indicative and Subjunctive. *Analysis* 48: 113–122.

Edgington, D. (1986) Do Conditionals have Truth-Conditions? *Conditionals*. Ed. F. Jackson. Oxford, Oxford University Press: 176–201.

— (2005).Ramsey's Legacy on Conditionals and Truth. *Ramsey's Legacy*. Ed. H. Lillehammer and D. H. Mellor. Oxford, Oxford University Press: 37–52.

— (2009) Conditionals, truth and assertion. *Minds, Ethics, and Conditionals: Themes From the Philosophy of Frank Jackson*. Ed. I. Ravenscroft. Oxford: Oxford University Press (2009): 283–310.

Feller, W. (1957). *An Introduction to Probability Theory and Its Applications*. 2nd Edition. New York, Wiley.

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

Gibbard, A. and W. L. Harper (1978). Counterfactuals and Two Kinds of Expected Utility. *Decision, Probability, and Utility*. Ed. P. Gärdenfors and N.-E. Sahlin. Cambridge, Cambridge University Press: 341–378

Gibbard, A. (1981) Two Recent Theories of Conditionals, *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Ed. W. L. Harper, R. C. Stalnaker and G. Pearce. Dordrecht: Reidel: 211–248.

Grice, H. P. (1975). Logic and Conversation. *Conditionals*. Ed. F. Jackson. Oxford, Oxford University Press: 155–175.

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

Jackson, F. (1990) Classifying Conditionals. *Analysis* 50(2): 134–147.

Jeffrey, R. C. (1965) *The Logic of Decision*. New York, McGraw-Hill.

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

Lewis, D. (1973) *Counterfactuals*. Oxford, Blackwell.

— (1976) Probabilities of Conditionals and Conditional Probabilities. *Philosophical Review* 85: 297–315.

McDermott, M. (1996) On the Truth Conditions of Certain If-Sentences. *Philosophical Review* 105: 1–37.

Mellor, D. H. (1993) How to Believe a Conditional. *Journal of Philosophy* 90: 233–248.

–- (2000) The Semantics and Ontology of Dispositions. *Mind* 109: 757–780.

–- (2005) What Does Subjective Decision Theory Tell Us? *Ramsey's Legacy*. Ed. H. Lillehammer and D. H. Mellor. Oxford, Oxford University Press: 137–148.

Milne, P. (1997) Bruno de Finetti and the Logic of Conditional Events, *British Journal for the Philosophy of Science* 48: 195–232.

Priest, G., F. Berto and Z. Weber (2018). Dialetheism. *The Stanford Encyclopedia of Philosophy*. Ed. E. N. Zalta. URL: <https://plato.stanford.edu/archives/fall2018/entries/dialetheism/>.

Ramsey, F. P. (1926). Truth and Probability. *F. P. Ramsey: Philosophical Papers*. Ed. D. H. Mellor. Cambridge, Cambridge University Press: 52–109.

Stalnaker, R. C. (1968). A Theory of Conditionals: Studies in Logical Theory, *American Philosophical Quarterly* (Monograph Series, 2). Oxford: Blackwell: 98–112.

— (1984) *Inquiry*. Cambridge, Mass., MIT Press.

Tooley, M. 1997. Time, Tense, and Causation. Oxford: Clarendon Press.