Received: 10 January 2020 | Revised: 20 April 2020 | Accepted: 1 June 2020

DOI: 10.1111/mila.12322

SUBMITTED ARTICLE

WILEY

Conditionals: Truth, safety, and success

Hugh Mellor[1†] | **Richard Bradley**[2]

[1]Faculty of Philosophy, Cambridge University, Cambridge, UK

[2]Department of Philosophy, Logic and Scientific Method, London School of Economics and Political Science, London, UK

Correspondence
Richard Bradley, London School of Economics and Political Science, Houghton Street London WC2A 2AE UK.
Email: r.bradley@lse.ac.uk

Whether I take some action that aims at desired consequence C depends on whether or not I take it to be true that if I so act, I will bring C about and that if I do not, I will fail to. And the action will succeed if and only if my beliefs are true. We argue that two theses follow: (I) To believe a conditional is to be disposed to infer its consequent from the truth of its antecedent, and (II) The conditional is true iff the inference would not make a true belief in the antecedent cause a false belief in the consequent.

KEYWORDS
conditionals, inferential dispositions, suppositions, actions, truth values

1 | INTRODUCTION

Whether conditionals have truth values and, if so, what they are, are still matters of controversy. The principal answers, for the case when "P" and "Q" are unconditional, true or false, and logically independent, include that:

† Professor Hugh Mellor, the primary author of this article, died on June 21, 2020, from complications arising from lymphoma. Hugh was an outstanding philosopher, best known perhaps for his work on the metaphysics of time, chance and causation. He was also instrumental in publicising the work of Frank Ramsey, whose pragmatist theory of belief greatly influenced Hugh. At the time of his death Hugh was an Emeritus Professor in the Faculty of Philosophy of Cambridge University, an institution with which he was associated for much of his academic career. This is the last article that Hugh wrote, and in it he develops a number of the recurring themes of his rich philosophical work: the dispositional theory of belief, the role of conditionals in our thinking, the causal theory of action and the relationship between all three.

A. "If P then Q" has no truth value (Adams, 1975; Edgington, 1986; Gibbard, 1981—for indicative conditionals);
B. "If P then Q" is true when "P" is true, if and only if "Q" is true; and when "P" is false either:
   1. Has either a third or no truth value (Belnap, 1970; Bradley, 2002; Edgington, 2005; McDermott, 1996; Milne, 1997); or
   2. Is true (Jackson, 1990; Lewis, 1976—for indicative conditionals); or
   3. 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 ordering of P-worlds (for instance, by how closely they resemble our ¬P-world), and false if it is not (Lewis, 1973; Stalnaker, 1984; Kratzer, 2012).

In this study, we derive a different answer to this question from the two-part theory that says that:

I. To *believe* "If P then Q" is to have a disposition, D(P > Q), to infer "Q" from "P", and
II. "If P then Q" is *true* if and only if in the circumstances the inference to "Q" from "P" is *safe*: That is, if and only if possession of the disposition D(P > Q) would not in the circumstances make a true belief in "P" cause a false belief in "Q".

The first thesis combines the "suppositional" view, that to believe "If P then Q" is to believe "Q" under the supposition that "P" is true (Adams, 1975; Gärdenfors, 1986; Edgington, 2009), with the view that to believe "Q" under the supposition that "P" is to be disposed to infer "Q" from "P": That is, the fact that it is a state of mind which makes believing "P" cause us to believe "Q" (Mellor, 1993; Stalnaker, 1984). For example, to believe that "If I take a walk, I'll get wet" is to be disposed to infer that I will get wet, a disposition triggered by my coming to believe in an appropriate way that I will take a walk, for example, in a way that does not involve my putting on a raincoat.[2]

The second thesis says that a conditional is true if and only if the inference to its consequent from its antecedent is safe. An inference from P to Q is safe in some circumstance if and only if in that circumstance Q is true supposing that P is. When an inference is safe in some circumstance, the disposition D(P > Q) would not in the circumstance make a true belief in "P" cause a false belief in "Q". In this case, we will say that the inferential disposition too is safe.

When "P" and "Q" are logically independent, the safety of the inference from "P" to "Q" is contingent on the circumstances. For example, the inference from my taking a walk to my getting wet is safe in circumstances when the weather is foul, but not when it is good. This is why, on claim (II), the sentence "If I take a walk, I'll get wet" is only contingently true or false. In contrast, an inference is valid if and only if it is safe in *all* circumstances: that is, if it is *necessarily*, and not just contingently, truth-preserving.

Our core argument for claims (I) and (II) is that they, unlike theories of type A–B3, can account for the role of conditionals in decision making. We start in Section 2 by saying how we take conditional beliefs to affect what action we perform as a means to an end. Suppose, for example, that I am deciding whether or not to pay £1 (P) to get a newspaper (Q). Then whether

[1]"Accessible" in the sense of Lewis (1973, section 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, Berto, & Weber, 2018), in no accessible world is any proposition both true and false.

[2]There are complications here that arise from whether the conditional is indicative or subjunctive, which are discussed in Section 5.2.

I make "P" true by paying or false by not paying depends, *inter alia*, on whether I believe both "If P then Q" and "If ¬P then ¬Q": That is, given (I), on whether I am 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 Section 3, we infer from (II) that for either of these actions (of making "P" true or making "P" false) to *succeed* as a sufficient means to its end (of 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), the fact 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 Section 4, the requirement cannot be met by any of the theories A–B3 listed earlier.

In Section 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 Section 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.

Finally, a note about the limits to what we do in this study. Although we contrast our theory of conditionals with several other prominent ones, we do not attempt to discuss all the views of conditionals to be found in the very large literature in philosophy, linguistics, and psychology (see Edgington, 1995; von Fintel, 2011; and Evans & Over, 2004 for good surveys in each of these fields). Nor do we do full justice to those theories that we do discuss; in particular, since the focus of the study is the role of conditionals in decision-making and the implications of this for their truth values, we do not attempt a full evaluation of these theories against evidence concerning other uses of conditionals. This latter task is important, but beyond the scope of this primarily theoretical study.

2 | ACTING ON CONDITIONALS

For present purposes, we need to only consider how our actions are affected by conditionals like "If P then Q" of which we are *certain*, that is, where we are disposed to *fully* believe "Q" if we believe "P". The decision theories of Ramsey (1926), Jeffrey (1965), and others 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 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 am usually quite sure I will get them if I pay for them and will not if I do not.

Although the theory we present generalises in a natural way to uncertainty, 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 how our conditional beliefs *will* affect our actions (Mellor, 2005), or prescriptively, as saying how they *should* affect them (Joyce, 1999; Bradley, 2017), perhaps because if they did not we would be irrational, which it is assumed we should not be. As for present purposes either reading will do, prescriptivists should in what follows read "will" as "should" where appropriate.

TABLE 1 Causes of action

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

Suppose then, in a more adaptable example, I have to decide, when booking a long-haul flight, whether to pay extra (E) to fly first 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 are as follows:

1. E&F: I pay extra and fly first;
2. E&¬F: I pay extra but do not fly first;
3. ¬E&¬F: I do not pay extra and do not fly first; and
4. ¬E&F: I do not pay extra but fly first 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 first, that is, 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 first and if I do not I *would not*, that is, the fact 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 either make "E" true if I think it *is* worth paying the extra in order to fly first, or make "E" false if I think it is *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.[3] This is displayed in Table 1, where X is whatever else it takes to book my flight.

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, that is, the fact that my belief in each of these conditionals is independent of whether it is counter-actual (has a false antecedent) or actual (has a true one).

3 | SUCCESS AND TRUTH

The alternative actions in Table 1 will *succeed*, by causing their respective ends, if I will fly first if I pay extra and would not if I do not. If they do succeed, then by our thesis (I), the inferences the dispositions $D(E > F)$ and $D(\neg E > \neg F)$ dispose us to make, from "E" to "F"; and from "¬E" to "¬F", will be safe: That is, $D(E > F)$ and $D(\neg E > \neg F)$ would not make a true belief in "E"

[3]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 first: 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 2 Truth Conditions of Causes

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

cause a false belief in "F" or a true belief in "¬E" cause a false belief in "¬F". And that, our thesis (II) says, is what makes those conditionals true, as is displayed in Table 2.

The upshot of this analysis is that the truth of these two conditionals is sufficient to explain why I will succeed in achieving my aim of flying first if and only if I pay extra and hence why this action is the correct one to choose (just as my believing them explains why I will—or should[4]—choose to pay more). Such an explanation cannot be offered by any theory (such as those of type A or B2), which denies that both conditionals can be true. A similar difficulty is faced by any theory (such as that of Lewis, 1973), which allows that when "E" is false, then neither "If E then F" nor "If E then ¬F" may be true. But if neither is true, then we cannot say whether or not it is correct to make "E" true.

That both can be true is an immediate consequence in our theory of the fact that it entails that exactly one of "If E then F" and "If E then ¬F" will be true. This is so because the inference from "E" to "F" is safe if and only if that from "E" to "¬F" is unsafe. So, the law of conditional excluded middle—that "(If E then F) or (If E then ¬F)" is logically true—is an implication of our theory. This law, particularly applied to counter-actual conditionals, is however contentious. One objection 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 second takes the *possibility* that "F" would be false if "E" was true to show that the counter-actual conditional "If E then F" *is* false. We respond to each objection 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 could not 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 cannot be.[5] And in whatever single E-world my paying extra *would* take me to, either I will fly first in it or I would not: So either "F" will be true in that world or "¬F" will be.

But if "F" *is* true in that world, that is, if "F" *would* be true supposing "E" *were*, 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 is actual and when it is counter-actual, as Table 2 shows; and so, for the same reason, can "If ¬E then ¬F".

[4]On prescriptive readings of decision theories, see Section 2.

[5]See footnote 1.

3.2 | "Might not" versus "would"

The mere fact that "F" *could* be false if "E" was true does not show that it *would* be false, nor hence that the counter-actual "If E then F" *is* false. All it shows is, as we say in Section 1, that "If E then F" is only contingently true. But if the merely logical possibility of "F" being false if "E" was true would not refute our counter-actual conditional, perhaps a *physical* possibility will. In other words, if "F" *might not* be true if "E" was true because there is a non-zero *chance* of its being false, might that not falsify the counter-actual "If E then F"?

Many philosophers have thought it would. Lewis (1973, p. 2) even makes the incompatibility of "might not" and "would" conditionals definitional, thus taking, in our example:

If I paid extra, I might not fly first = df ¬(If I paid extra, I would fly first)

If I paid extra, I would fly first = df ¬(If I paid extra, I might not fly first)

So, suppose that, unknown to me, I *might not* fly first if I paid extra, because there is a real chance of my aircraft being grounded by bad weather or engine failure. Then while I *might* still fly first if I paid extra, I *might also not* fly first: a conjunction that for Lewis falsifies *both* the counter-actuals "If I paid extra I *would* fly first" and "If I paid extra I *would not* fly first".[6]

Our reason for rejecting Lewis' thesis, and hence the falsehood of both "If I paid extra I *would not* fly first" and "If I paid extra I *would* fly first" is the far simpler and more general one given in Section 3.1: Paying extra could not take me to a non-actual world in which I neither fly first nor do not fly first. And while we do not deny the conflict between "I might not fly first if I paid extra" and "I would fly first if I paid extra", our theory makes it not logical but epistemological: I cannot be *sure* I would fly first if I paid extra if I *know* there is a chance that I would not. For then my coming to believe "E" would not cause me to *fully* believe "F": It would only cause me to have a credence (a probability measure of my degree of belief) less than 1. But none of this alters the fact that any counter-actual world in which "E" is true that is an F-world, as it may well be, makes the counter-actual "If E then F" true, just as it would be if it was actual and both "E" and "F" were actually true.

4 | THEORIES A–B3

We showed in Section 2 how conditionals we fully believe affect what we do as a means to an end, and in Section 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 Section 1 can explain these results.

Firstly, as we argued earlier, any type A theory'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. This is why it cannot explain why the correct action to choose is the one with the preferred end.

Similarly a type B1 theory'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 the truth of "E" cause "F" to be true or the falsity of "E" to cause "F" to be false. So,

[6]Lewis later (1979) admits that this may fail to follow, but only when not flying first if I paid extra would be a "quasi-miracle", that is, both improbable and as remarkably systematic as a "monkey at a typewriter [producing] a 950 page dissertation" (p. 60). For since a quasi-miraculous E&¬F-world would be less similar to our ¬E world than any E&F-world, Lewis's possible-world semantics would then let the counter-actual "If E then F" be true. However, this exception, criticised anyway by Hawthorne (2005) and others, will clearly not save our example: For not only cannot the chances of my not flying first and of my flying first *both* be small enough to be quasi-miraculous, *neither* may be, since they may both be closer to 0.5 that to 0 or 1.

such a theory cannot explain what makes all the available actions successful in achieving their ends: It can only explain the success of the one actually chosen.

The trouble with B2's truth-functional "material" conditionals is not that they are *not* true when they are counter-actual but that they are *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 Section 3.1 they cannot 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 cannot be disposed to believe both "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" is not.

B3 theories, which say that when "E" is false, "If E then F" will be true in our ¬E-world if and only if "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 twofold. Firstly, when "F" is true in some but not all highest-ranked E-worlds, such theories make both "If E then F" and "If E then ¬F" false. But, as we argued earlier, if neither is true, then we cannot say whether it is correct to make "E" true or not, because there is no fact of the matter as to what we would achieve by doing so.[7]

Secondly, what *makes* "If E then F" true when "E" is false is not any *overall* ranking of non-actual E-worlds, for example, their overall similarity to our ¬E world (Lewis, 1973, chapter 2.3). To show this, we recall first that most of the conditionals we act on are contingent. 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 enables aircraft to fly, the conjunction of which we will call "Y". It is also contingent on many variable local facts—the airline's pricing policy, the availability of first-class seats, the plane's taking off, and so forth—the conjunction of which we will call "Z". And what determines whether a non-actual E-world is also an F-world is not 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 first (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 do not 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 how their truth makes these actions succeed, we must now show how the theory applies to other kinds of conditionals, including those that have only an indirect, if any, relation to action.

[7]This objection does not apply to Stalnaker (1968) who assumes that worlds are strictly ordered. Our second objection does apply to Stalnaker.

5.1 | Third-person and causal conditionals

The conditionals we act on are future-referring first-person ones like "If I pay extra, I'll fly first" 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 is true of this conditional is also true of its third-person 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.

The inference we make in this case concerns a particular person. But the general thought underlying it can be expressed by an impersonal causal conditional such as

If you run and are fit, you will finish,

which we also believe whether or not any particular fit person runs, since it is their fitness that makes it safe to infer they finishing from their running.

Of course, many causal conditionals make no explicit reference to an action in the antecedent; for example:

"If nurses' wages rise, more people will train to be nurses"

Nonetheless such a conditional can also serve as a basis for action, for its truth (if it is true) is sufficient for raising wages to succeed as a means of getting more people training to be nurses. So insofar as causal conditionals can be acted upon, they must have truth values that are independent of those of their antecedents. And this is a feature that is explained by our theory and not those of type A–B3.[8]

5.2 | Indicative and subjunctive conditionals

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

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

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

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

[8]No particular theory of causation is assumed here or anywhere else in our paper. It is true that the connection between action and causation is closest on interventionist theories of causation, but our argument does not depend on such a theory.

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

C1: If Anna runs, she will finish

is admittedly contentious, being accepted by some philosophers (Bradley, 2017; Gibbard & Harper, 1978; Joyce, 1999) but rejected by others (Bennett, 1988; Dudman, 1988). Here, while we need not 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 implying that Anna does not run.

The problem which simultaneous believing C2 and disbelieving 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.[9] 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 is 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 would not 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 would not run unless she is fit enough to finish; what makes the causal and counter-actual C3 false is that she is not that fit.

5.3 | Past-referring conditionals

Suppose first that Anna's fitness does make the future-referring C1, "If Anna runs, she will 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 Section 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.

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.

[9]See footnote 1.

If Oswald had not killed Kennedy someone else would have

if we believe Oswald had backup, since we know that only then would the disposition D (Oswald did not kill Kennedy > someone else did) be safe. However, since we know that Kennedy *was* killed, we do believe the evidential and actual

If Oswald did not kill Kennedy someone else did

whether or not we believe he had backup.[10]

5.4 | Centring and missing-link conditionals

Another consequence of our theory, as of those of type B1-3, is the so-called "centring" principle (Lewis, 1973, p. 14). Centring says that, for all "P" and "Q", "If P then Q" is true if both "P" and "Q" are. This is uncontentious when P and Q are causally linked, as my paying extra and flying first are by the facts Y&Z cited in Section 4, and Anna's running and finishing are by her fitness. The principle is more contentious when they are not so related, as in so-called "missing-link" conditionals, such as "If London is large, water is wet". Although an assertion of this sentence sounds odd, our theory makes it true. This is because, since London *is* large and water *is* wet, the inferential disposition it expresses is in fact safe.

In this aspect, our theory differs from other theories relating the truth of a conditional to properties of the inference from the conditional's antecedent of the conditional to its consequent. Let us call any theory "inferentialist" if it says that a conditional is true if and only if its consequent can permissibly be inferred from its antecedent. Different specifications of what inferences are permissible will then generate different truth-conditions for conditionals. Our theory, which in effect says that the inference is permissible if and only if it is safe, lies at one end of a spectrum. At the other end is the theory of strict conditionals, which says that only deductively valid inferences (ones that are necessarily safe) are permissible and hence that conditionals are true only if their antecedents entail their consequents (Gillies, 2009). Recent inferentialist theories occupy a position lying between these two, requiring that the antecedent and consequent be linked by an acceptable argument of some kind—inductive, abductive or deductive—perhaps drawing on implicit background knowledge (Krzyżanowska, Wenmackers, & Douven, 2013; Douven, 2017). Of course, to get truth-values a precise account of the valid non-deductive inference is required, but these details are not important here.

Inferentialist theories of this intermediate kind give a semantic explanation for the oddness of the assertion of missing-link conditionals. For example, the fact that they are false because inferences from their antecedents to their consequents do not meet the theories' standards of acceptability. But we reject these theories too for the same reason as we reject other theories

[10]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 did not kill Kennedy someone else did", by disposing us to infer its consequent from its antecedent: A disposition that is 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.

that allow both “E then F” and “E then ¬F” to be false: They cannot explain the relation between the truth of a conditional and the success of actions that make its antecedent true.

Worse, on such accounts a conditional can be false even when it is safe, that is, when making “E” true suffices for the truth of “F”. Thus, suppose in our flying example that, oddly enough, I *would* get to fly first class even if I *did not* pay extra. Then on our account “If I do not pay extra, I will fly first” is true, and this fact explains why I succeed in achieving my aim of flying first, despite not paying extra. But as “I do not pay extra” does not plausibly justify “I fly first”, the moderate inferentialist must deem the conditional that I successfully act upon to be false.

Nonetheless, there is something right about the inferentialist's observation that we do not assert conditionals that are not reliably safe, because we require something more than a purely accidental connection between the truth value of their consequents and that of their antecedents. But this assertibility constraint is pragmatic—a conversational implicature perhaps—rather than semantic. An assertion of “If London is large, water is wet” sounds odd, for instance, because it tacitly implies something false, namely that we believe London's largeness makes water wet, which we do not. But this no more shows that “If London is large, water is wet” is untrue than, as we saw in Section 3.2, the epistemic impropriety of conjoining “might not” and “would” counter-actuals shows that the former's truth falsifies the latter.

5.5 | Non-contingent conditionals

Although non-contingent conditionals often have little connection to decision-making, our theory as expressed by theses (I) and (II) still applies to them.

(i) If a contingent “P” deductively entails a contingent “Q”, the inference to “Q” from “P”, is necessarily safe and that to “¬Q” necessarily unsafe. This 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 Section 5.4's “If London is 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 cannot 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, for example, of *reductio* proofs of the form “If P then ¬P; therefore ¬P”, like:

“If there is a greatest prime number, p_N, then $p_1p_2...p_N + 1$ 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?[11] The answer is that, since believing “P” could not 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: For example, to revert to our flying example,

"If I fly Virgin, I'll fly first 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", that is, 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)) 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 would not 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", that is, "If I'll fly first if I pay extra, I'll fly Virgin". The disposition expressed by this conditional, 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 Section 5 how the truth values of all kinds of conditionals correspond credibly with the safety values of the inferential dispositions they express. But this is not 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, section 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 is *actual*, that is, when "P" is actually true, and when it is *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.

However, when "P" is *false* in w, so that w is *not* a P-world w_E, "If P then Q" will be counteractual 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 inferences they express.

ACKNOWLEDGMENTS

The authors are grateful for the comments on earlier drafts by James Cargile, Dorothy Edgington, Alan Hájek, the audience of the Serious Metaphysics Group in Cambridge and the two anonymous referees for *Mind & Language*.

REFERENCES

Adams, E. (1975). *The logic of conditionals*. Dordrecht: Reidel.

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge.

Belnap, N. D. (1970). Conditional assertion and restricted quantification. *Noûs*, *4*(1), 1–12. https://doi.org/10.2307/2214285

Bennett, J. (1988). Farewell to the phlogiston theory of conditionals. *Mind*, *97*(388), 509–527. https://doi.org/10.1093/mind/XCVII.388.509

Bradley, R. (2002). Indicative conditionals. *Erkenntnis*, *56*(3), 345–378. https://doi.org/10.1023/A:1016331903927

Bradley, R. (2012). Multidimensional possible-world semantics for conditionals. *Philosophical Review*, *121*(4), 539. https://doi.org/10.1215/00318108-1630921

Bradley, R. (2017). *Decision theory with a human face*. Cambridge: Cambridge University Press.

Douven, I. (2017). How to account for the oddness of missing-link conditionals. *Synthese*, *194*(5), 1541–1554. https://doi.org/10.1007/s11229-007-9214-5

Dudman, V. H. (1988). Indicative and subjunctive. *Analysis*, *48*(3), 113–122. https://doi.org/10.1093/analys/48.3.113a

Edgington, D. (1986). Do conditionals have truth-conditions? In F. Jackson (Ed.), *Conditionals* (pp. 176–201). Oxford: Oxford University Press.

Edgington, D. (1995). On conditionals. *Mind*, *104*(414), 235–329. https://doi.org/10.1093/mind/104.414.235

Edgington, D. (2005). Ramsey's legacy on conditionals and truth. In H. Lillehammer & D. H. Mellor (Eds.), *Ramsey's legacy* (pp. 37–52). Oxford: Oxford University Press.

Edgington, D. (2009). Conditionals, truth and assertion. In I. Ravenscroft (Ed.), *Minds, ethics, and conditionals: Themes from the philosophy of Frank Jackson* (pp. 283–310). Oxford: Oxford University Press.

Evans, J. S. B. T, & Over, D. E. (2004). *If*, Oxford: Oxford University Press.

Gärdenfors, P. (1986). Belief revisions and the Ramsey test for conditionals. *The Philosophical Review*, *95*(1), 81–93. https://doi.org/10.2307/2185133

Gibbard, A. (1981). Two recent theories of conditionals. In W. L. Harper, R. C. Stalnaker & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance, and time* (pp. 211–248). Dordrecht: Reidel.

Gibbard, A. & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In P. Gärdenfors & N.-E. Sahlin (Eds.), *Decision, probability, and utility* (pp. 341–378). Cambridge: Cambridge University Press.

Gillies, A. S. (2009). On truth-conditions for if (but not quite only if). *Philosophical Review*, *118*(3), 325. https://doi.org/10.1215/00318108-2009-002

Hawthorne, J. (2005). Chance and counterfactuals 1. *Philosophy and Phenomenological Research*, *70*(2), 396–405. https://doi.org/10.1111/j.1933-1592.2005.tb00534.x

Jackson, F. (1990). Classifying conditionals. *Analysis*, *50*(2), 134–147. https://doi.org/10.1093/analys/50.2.134

Jeffrey, R. C. (1965). *The logic of decision*. New York, NY: McGraw-Hill.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge: Cambridge University Press.

Kratzer, A. (2012). *Modals and conditionals*, Oxford: Oxford University Press.

Krzyżanowska, K., Wenmackers, S. & Douven, I. (2013). Inferential conditionals and evidentiality. *Journal of Logic, Language and Information*, *22*(3), 315–334. https://doi.org/10.1007/s10849-013-9178-4

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

Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review*, *85*(3), 297–315. https://doi.org/10.2307/2184045

McDermott, M. (1996). On the truth conditions of certain "if"-sentences. *The Philosophical Review*, *105*(1), 1–37. https://doi.org/10.2307/2185762

Mellor, D. H. (1993). How to believe a conditional. *The Journal of Philosophy*, *90*(5), 233–248. https://doi.org/10.2307/2940911

Mellor, D. H. (2000). The semantics and ontology of dispositions. *Mind*, *109*(436), 757–780. https://doi.org/10.1093/mind/109.436.757

Mellor, D. H. (2005). What does subjective decision theory tell us? In H. Lillehammer & D. H. Mellor (Eds.), *Ramsey's legacy* (pp. 137–148). Oxford: Oxford University Press.

Milne, P. (1997). Bruno de Finetti and the logic of conditional events. *The British Journal for the Philosophy of Science*, *48*(2), 195–232. https://doi.org/10.1093/bjps/48.2.195

Priest, G., Berto, F. & Weber, Z. (2018). Dialetheism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Retrieved from. https://plato.stanford.edu/archives/fall2018/entries/dialetheism

Ramsey, F. P. (1926). Truth and probability. In D. H. Mellor (Ed.), *F. P. Ramsey: Philosophical papers* (pp. 52–109). Cambridge: Cambridge University Press.

Stalnaker, R. C. (1968). A theory of conditionals. In *Studies in logical theory* American Philosophical Quarterly Monograph Series (Vol. 2, pp. 98–112). Oxford: Blackwell.

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

von Fintel, K. (2011). Conditionals. In K. Von Heusinger, C. Maienborn & P. Portner (Eds.), *Semantics: An international handbook of meaning vol. 2 (Handbücher zur sprach- und kommunikationswissenschaft 33.2)* (pp. 1515–1538). Berlin/Boston: de Gruyter Mouton.

How to cite this article: Mellor, H. & Bradley, R. (2022). Conditionals: Truth, safety, and success. *Mind & Language*, *37*(2), 194–207. https://doi.org/10.1111/mila.12322