

This is an electronic version of a paper to be published in Pacific Philosophical Quarterly.

## **Possible World Semantics and True-True Counterfactuals**

**Lee Walters**

The standard semantics for counterfactuals ensures that any counterfactual with a true antecedent and true consequent is itself true. There have been many recent attempts to amend the standard semantics to avoid this result. I show that these proposals invalidate a number of further principles of the standard logic of counterfactuals. The case against the automatic truth of counterfactuals with true components does not extend to these further principles, however, so it is not clear that rejecting the latter should be a consequence of rejecting the former. Instead I consider how one might defuse putative counterexamples to the truth of true-true counterfactuals.

One allegedly undesirable feature of the standard Stalnaker-Lewis possible world semantics for counterfactuals is that conditionals with true components are themselves trivially true (§1). Lewis proposes a semantics which lacks this feature, but this too has been deemed unsatisfactory (§2). As a result, several authors have tried to revise the standard semantics so as to avoid the automatic truth of counterfactuals with true components (§4-§7). I note, however, (§3) that as well as making counterfactuals with true components automatically true, the standard semantics validates a range of plausible and popular principles. Then (§4-§7) I show how each modification of the standard account considered here requires a number of these further principles to be rejected. Such accounts, then, are logically revisionary in ways that many opponents of the automatic truth of true-true counterfactuals find objectionable. I close by

highlighting the lessons to be learnt from the discussion of these semantic proposals (§8), and by suggesting what the defender of the trivial truth of true-true counterfactuals can say in response to putative counterexamples (§9).

### 1. The Putative Problem

On the standard semantics for counterfactuals (Stalnaker 1987; Lewis 1986), a counterfactual conditional,  $A \Box \rightarrow C$ , is true at  $w$  iff (i) there is an  $(A \wedge C)$ -world closer to  $w$  than every  $(A \wedge \sim C)$ -world, where closeness is a matter of similarity, or else (ii) there are no  $A$ -worlds. This semantics validates

Conjunction Conditionalization:  $(A \wedge C) \supset (A \Box \rightarrow C)$

because Stalnaker and Lewis also embrace the following constraint

Strong Centring: Any world is more similar to itself than any other world is to it.

Many authors, though, reject Conjunction Conditionalization on the basis of putative counterexamples, such as McDermott's (2007): a coin is to be tossed twice; before it is tossed I bet that it will come up heads both times; it does and I win. Now consider

- (1) If at least one head had come up, I would have won.

(1) is true on the standard account since it follows from Conjunction Conditionalization – at least one head did come up and I did win. McDermott claims, however, that intuitively (1) is false and so Conjunction Conditionalization is invalid. McDermott concludes that the standard semantics – in particular the combination of (i) and Strong Centring – is to be rejected.

## **2. Weak Centring**

The alleged invalidity of Conjunction Conditionalization has led to the development of rival possible world semantics which lack this consequence. In order to avoid Conjunction Conditionalization, Lewis (1986: 22) himself considers amending the standard semantics by replacing Strong Centring with

Weak Centring: Any world is amongst the most similar worlds to itself,

whilst maintaining (i) and (ii). Bennett (1974: 387) claims that this manoeuvre is ineffective since no world can be as similar to  $w$  as  $w$  itself is. Cogburn and Roland (2013: 252), McGlynn (2012: 277), and Penczek (1997: 80-81) agree. Given that these authors all wish to maintain Modus Ponens which, given (i), is secured by Weak Centring, these authors conclude that the core of the standard semantics, (i), is to be rejected. In the light of this result, they then provide their own possible world semantics which invalidate Conjunction Conditionalization.

Bennett's thought above, however, is incorrect. Although Strong Centring might be non-negotiable given some intuitive, but not purely qualitative, notion of similarity, Lewis (1979) and Stalnaker (1987) admit, in response to objections from Bennett (1974) and Fine (1975), that

similarity here is a technical notion and not what we might think of as intuitive overall similarity.<sup>i</sup> As Stalnaker puts it, Bennett's and Fine's objections "show decisively that the intuitive notion of overall similarity between possible worlds ... is not the [notion] that is relevant to the interpretation of counterfactual conditionals" (1987: 127). Moreover, such "[a]n account of the respects of similarity that are relevant to selection [of possible worlds] might say that some respects of similarity count for nothing at all, and so should be ignored" (Stalnaker 1987: 128). Such a measure

may nevertheless be a relation of overall similarity – not because it is likely to guide our explicit judgments of similarity, but rather because it is a resultant, under some system of weights or priorities, of a multitude of relations of similarity in particular respects (Lewis 1979: 43).

As a result, it is not clear that a world is *more* similar to itself than any other world is. So, given that Cogburn and Roland, McGlynn, and Penczek all motivate their proposals to invalidate Conjunction Conditionalization on the basis of Bennett's inaccurate complaint, their accounts are under supported.<sup>ii</sup> That is, for all Bennett has said, Conjunction Conditionalization can be avoided by adopting Lewis's suggestion of maintaining (i) but retreating from Strong Centring to Weak Centring.<sup>iii</sup>

McDermott (2007: 334), noting the failure of Bennett's objection, argues that nevertheless the combination of (i) and Weak Centring is not an option for those who reject (1). McDermott's

case against this package can be put as follows. In McDermott's scenario described above, as well as rejecting (1), McDermott also endorses

(2) If I had bet on two heads, two heads would have come up,

just as we endorse

(3) If I had not bet on two heads, two heads would have come up.

(3) is an 'irrelevant semifactual': it is a semifactual (a counterfactual with a true consequent and false antecedent) where the obtaining of the antecedent is irrelevant to the obtaining of the consequent. And it is a commitment of our thinking about counterfactuals that such irrelevant semifactuals are true. The reason for this is that when considering counterfactuals we hold fixed the effects of causal chains that are independent of the obtaining of the antecedent. So given that two heads came up and that the result of the coin toss is independent of my betting behaviour, (3) is true. Similar reasoning supports (2).

Some, like Phillips (2007), would reject this line of thinking since they reject that irrelevant semifactuals are true in indeterministic contexts. Nevertheless, this line of thinking is widely accepted by advocates and critics of Conjunction Conditionalization alike (Bennett 2003: §9, Edgington 2004, Kvart 1986, McDermott 2007, Noordhof 2004, Pollock 1976: 26, Schaffer 2004, and Walters 2009). Moreover, Lewis (1979: 48 – see his discussion of Morgenbesser's coin) accepts the truth of irrelevant semifactuals, as do Cogburn and Roland (2013), McGlynn

(2012), and, more tentatively, Penczek (1997). So these authors would accept (2). In any case, we can stipulate that we are to think of McDermott's situation as deterministic. In such a context, (2) seems secure.

Now if Lewis's retreat from Strong Centring to merely Weak Centring is to account for the falsity of (1), some world,  $w^*$ , at which only one head comes up and I lose (that is, where I don't bet on exactly one head coming up) must be as close to the actual world, @, as @ is to itself. In particular  $w^*$  is a world where I still bet on two heads. That is, what in part underwrites the intuitive falsity of (1) is

(4) If at least one head had come up, I would have bet two heads.

Moreover, that (4) is true can be seen by noting that it is an irrelevant semifactual: I did bet two heads and my betting behaviour is independent of the result of the coin toss.

So  $w^*$  is a world where I bet on two heads (from (4)), and also a world where only one head comes up (from the falsity of (1)). But as  $w^*$  is as close to @ as @ is,  $w^*$  is sufficient to make (2) false given (i). So, given the truth of (2), the falsity of (1) cannot be accommodated by maintaining both (i) and Weak Centring. Now (i) only validates Modus Ponens in conjunction with Weak Centring, so (i) has to be rejected if Modus Ponens is to be upheld. The possible world semanticist, then, will have to look elsewhere to accommodate failures of Conjunction Conditionalization. In §4-§7 I consider some alternative proposals. But first, let's examine some further consequences of the standard account.

### 3. The Logic of Counterfactuals

As well as Conjunction Conditionalization, the standard logic of counterfactuals validates the principles below. Indeed, if we remove Conjunction Conditionalization from the axiomatization of Lewis's official system, VC, we have a weaker logic, VW, which validates the principles below without also validating the allegedly problematic Conjunction Conditionalization

Modus Ponens:  $(A \Box \rightarrow C) \supset (A \supset C)$

Agglomeration:  $((A \Box \rightarrow B) \wedge (A \Box \rightarrow C)) \supset (A \Box \rightarrow (B \wedge C))$

Weakening the Consequent: if  $\vdash (B \supset C)$ , then  $(A \Box \rightarrow B) \supset (A \Box \rightarrow C)$

Disjunction:  $((A \Box \rightarrow C) \wedge (B \Box \rightarrow C)) \supset ((A \vee B) \Box \rightarrow C)$

SDA\*:  $((A \vee B) \Box \rightarrow C) \supset ((A \Box \rightarrow C) \vee (B \Box \rightarrow C))$

VLAS:  $((A \Box \rightarrow B) \wedge (A \Box \rightarrow C)) \supset ((A \wedge B) \Box \rightarrow C)$

Substitution:  $((A \Box \rightarrow B) \wedge (B \Box \rightarrow A) \wedge (A \Box \rightarrow C)) \supset (B \Box \rightarrow C)$

Limited Transitivity:  $((A \Box \rightarrow B) \wedge ((A \wedge B) \Box \rightarrow C)) \supset (A \Box \rightarrow C)$ .

As we shall see, the four accounts discussed below do not validate VW in virtue of not validating all of the above principles. But the accounts discussed below are attempts to *conservatively* revise the standard semantics. So, to the extent that they invalidate some of the above principles, they fail in this respect. Further, with the exception of Gundersen on Modus Ponens, none of the authors discussed below note that their semantics invalidate some of the above principles, let

alone provide arguments that the above principles are invalid. So, for all these authors have said, their semantics and the logics they validate are undermotivated.

The above notes a failure in the argumentative strategy of the authors who reject Conjunction Conditionalization, but this is not to say that the theorems of VW listed above are valid. So what can be said in their favour? Well, we should note that many who wish to reject Conjunction Conditionalization will wish to retain the above attractive looking principles, and so the semantics considered below will be unacceptable to them.<sup>iv</sup> Further, although the above principles are theorems of the standard possible world semantics, supplemented with Weak Centring, these principles are not tied to such a semantics. For instance, these principles are validated by Pollock's (1976: 42-43) cotenability semantics and Gärdenfors (1978) belief revision semantics, and they correspond to analogous claims about probabilistic entailment on Adams' (1975: 61) probabilistic account.

It is, however, difficult to provide a full defence of these principles as no appeal to examples can establish the validity of a theorem. And it is hard to know what to say to someone who does not find the above principles compelling; at some point we reach bedrock, and can only point to the plausibility of principles. I take it, however, that the first three principles are extremely intuitive and that most who reject Conjunction Conditionalization will want to do so without going to the extremes of rejecting these principles. For example, it is very difficult to see how Agglomeration could be false. Isn't it obvious that the truth of 'if I had gone to the party, Jasmine would have left' and 'if I had gone to the party, Priya would have left' licenses the claim that if I had gone to

the party, Jasmine and Priya would have left? With the exception of McGlynn, all of the authors below invalidate at least one of these three principles.

The two principles concerning disjunctive antecedents also seem compelling. Burgess (1981: 77) and Pollock (1976: 42-43) have Disjunction as an axiom, and Pruss, a critic of the standard account, also accepts Disjunction arguing that since

the subjunctive conditional says what would happen were some condition realized ... if some proposition would hold under one condition and would also hold under another condition, then it is hard to deny that it would hold under the disjunction of these two conditions (Pruss 2007: 33-34).

Apart from agreeing with these philosophers we should note that one reason for rejecting Disjunction does not undermine our arguments against the accounts below. Disjunction is suspect, one may think, because one way in which  $A \vee B$  can be true is when  $A \wedge B$  is true. So instead of endorsing Disjunction, one may only be prepared to endorse

Disjunction\*:  $((A \Box \rightarrow C) \wedge (B \Box \rightarrow C) \wedge ((A \wedge B) \Box \rightarrow C)) \supset ((A \vee B) \Box \rightarrow C)$ .

But all of the proposals below invalidate Disjunction\* as well as Disjunction, and so in what follows I'll concentrate only on the weaker principle.<sup>v</sup>

In addition, all the accounts below invalidate SDA\*. One might try to mitigate this cost by noting that we are often misled by counterfactuals with disjunctive antecedents, and so SDA\* might not be as innocent as it seems. For example, although

$$\text{SDA: } ((A \vee B) \Box \rightarrow C) \supset ((A \Box \rightarrow C) \wedge (B \Box \rightarrow C))$$

also looks good, it is invalid on the standard account. However, there are good reasons to reject SDA. First, SDA prevents us from substituting logical equivalents in the antecedents of conditionals: consider some true counterfactual,  $A \Box \rightarrow C$ . By substitution of logical equivalents we have  $(A \vee (A \wedge \sim C)) \Box \rightarrow C$ . SDA then allows us to conclude,  $(A \wedge \sim C) \Box \rightarrow C$ . But this is necessarily false when  $A$  and  $\sim C$  are compossible. Second, there are good counterexamples to SDA. To take McKay and van Inwagen's (1977: 355) example; from

- (5) If Spain had fought on either the Allied side or the Axis side, she would have fought on the Axis side

it does not follow that

- (6) If Spain had fought on the Allied side, she would have fought on the Axis side.

But note that this case against SDA does not extend to SDA\*. So even if we do think that we are misled by disjunctive antecedents in the case of SDA, this consideration does not tell against SDA\*. Of course, this observation does not itself establish SDA\*, but it does look good

All of the accounts below also invalidate the remaining three principles, VLAS, Substitution, and Limited Transitivity. The validity of these principles is, though, less obvious. So, is it not open to those who reject Conjunction Conditionalization to reject these principles? Indeed, I have argued elsewhere (Walters: 2009; and Walters and Williams: 2013) that Substitution and VLAS must go if we reject Conjunction Conditionalization. So perhaps those who reject Conjunction Conditionalization will be happy to pay this price. Moreover, counterexamples to Substitution and VLAS have been proposed by Ahmed (2011), Tichý (1978), and Tooley (2002).<sup>vi</sup> Nevertheless, we should note that Burgess (1981: 77) and Pollock (1976: 42-43) take VLAS as axiomatic, that VLAS follows from SDA\*<sup>vii</sup> (see the appendix to Walters and Williams: 2013), and that some opponents of Conjunction Conditionalization are committed to VLAS: ‘[t]here can be no doubt’ about the validity of VLAS (Bennett 2003: 332).

Given the controversial nature of Substitution and VLAS, I’ll focus on Limited Transitivity in what follows. Is it plausible, then, to reject Limited Transitivity? First, we should note that, as far as I know, no counterexamples to Limited Transitivity have been proposed in the literature.<sup>viii</sup> Second, Limited Transitivity is endorsed by many who reject Conjunction Conditionalization, (e.g. Bennett 2003; Gauker 2005: 248; and Lowe 1995). Finally, as well as being an intuitive and popular principle, Limited Transitivity plays an important explanatory role. The following principle is invalidated by the standard semantics and by all of the accounts below

Transitivity:  $((A \Box \rightarrow B) \wedge (B \Box \rightarrow C)) \supset (A \Box \rightarrow C)$ .<sup>ix</sup>

But Transitivity often looks very good, as it seems to underlie arguments like the following

- (7) If you had jumped from the roof, you would have broken some bones.
- (8) If you had broken some bones, you would have been in pain.
- (9) Therefore, if you had jumped from the roof, you would have been in pain.

The problem with Transitivity is that it also seems subject to counterexamples such as the following

- (10) If Smith had died, Jones would have stopped campaigning.
- (11) If Jones had stopped campaigning, Smith would have won the election.
- (12) Therefore, if Smith had died, Smith would have won the election!

Whilst (10) and (11) could well be true, (12) seems false, and so Transitivity is, it appears, invalid. Stalnaker and Lewis take the counterexamples to Transitivity at face value, as, it seems, do the authors discussed below.<sup>x</sup> But this means that such authors cannot appeal to Transitivity when explaining the seeming validity of (7)-(9). Instead, Stalnaker and Lewis note that although Transitivity is not valid, Limited Transitivity is. Moreover, instances of Transitivity seem compelling only when we are prepared to strengthen the antecedent of the second premise so as to deliver an argument which is an instance of Limited Transitivity. So, the thought goes, it is Limited Transitivity that underwrites (7)-(9), not Transitivity.

To see how this works, consider the required strengthening of the second premises of the instances of Transitivity above. In the first argument we are prepared to accept the conclusion because we are prepared to strengthen (8) to

- (13) If you had jumped from the roof and broken some bones, you would have been in pain.

As a result we have, given Limited Transitivity, a sound argument from (7) and (13) to (9). In the second argument, however, we are not prepared to accept the conclusion. The diagnosis is that in this case we are not prepared to strengthen (11) to

- (14) If Smith had died and Jones had stopped campaigning, Smith would have won the election.

In the envisaged scenario, (14) is false, and so Limited Transitivity does not license the conclusion (12). So whereas Stalnaker and Lewis can explain why the good instances of Transitivity seem good, and why the bad instances of Transitivity seem bad, the authors below cannot as the explanation crucially rests on the validity of Limited Transitivity which they fail to secure.

In what follows I show that all of the accounts below invalidate Limited Transitivity, SDA\*, and Disjunction\*, as well as each being subject to at least one further difficulty. Such attempts to invalidate Conjunction Conditionalization, then, will not hold universal appeal.

#### 4. Penczek

A natural *first* thought about counterfactuals is that their truth requires some sort of connection between the antecedent and the consequent. Moreover, many opponents of Conjunction Conditionalization

including Bennett (1974, pp. 386-388), Fine (1975, p. 453), and Bigelow (1976, p. 218), have tacitly argued on informal semantic grounds from what might be called the Connection Hypothesis: a necessary condition for the truth of a subjunctive conditional of the form  $A \square \rightarrow C$  is the existence of some sort of 'connection' ... between the propositions expressed by  $A$  and  $C$ . Mere concurrent or coincident truth, so the argument runs, is either not a connection at all or at least not one of the requisite sort. Hence, since [Conjunction Conditionalization] would secure the truth of subjunctive conditionals with true but unconnected antecedents and consequents, it should be reckoned invalid, the critics maintain, for exactly this reason (Butcher 1983: 71).

Penczek seems to endorse this objection to Conjunction Conditionalization when he writes that many putative counterexamples to Conjunction Conditionalization

exploit the fact that counterfactual statements, in virtue of their form, suggest a certain sort of connection (often, but not always, a causal connection) between antecedent and consequent, while in fact this connection may be lacking. That

counterfactuals with true components should automatically be true might thus be considered a weakness in Lewis's account (Penczek 1997: 80).

Penczek concludes from this that the mere truth of A and C should not count in favour of  $A \Box \rightarrow C$ . As a result, Penczek suggests that when assessing  $A \Box \rightarrow C$  we should consider situations in which A and C are both false. To see how Penczek does this, let's introduce a two-place connective,  $>$ , so that  $A > C$  has the truth conditions that the standard account attributes to  $A \Box \rightarrow C$ . We then use  $A > C$  to provide Penczek's semantics for counterfactuals:

$A \Box \rightarrow C$  is true at a world  $w$  iff either:

- a. there is no A-world, or else both
- b.  $A > C$ , and,
- c.  $(\sim A \wedge \sim C) > (A > C)$ .

When A and C are false, (c) is redundant, given that  $>$  obeys modus ponens, and so Penczek's truth condition matches the standard account's. But when at least one of A or C are true, we have to go to the closest  $(\sim A \wedge \sim C)$ -worlds, and assess whether or not  $A > C$  is true there, as well being true at the world of evaluation. As a result, Penczek's account predicts that (1) is false just as the opponent of Conjunction Conditionalization wants it to be. In particular, condition (c) is not met: in the closest worlds in which no heads come up and I lose the bet, 'at least one head  $>$  I win' is false, since amongst the closest worlds where at least one heads comes up will be worlds where I still bet two heads and only one heads comes up.

There is, however, a general challenge to those, like Penczek, who endorse the Connection Hypothesis, namely that it seems to be false for reasons independent of Conjunction Conditionalization. In particular, the Connection Hypothesis cannot account for the truth of irrelevant semifactuals such as (3) and ‘if I had scratched my nose, the coin (still) would have landed heads’. Such conditionals are true precisely because the consequent is true and the obtaining of the antecedent makes no difference to this. But given that such conditionals have true consequents, Penczek’s semantics requires that  $A > C$  be true at the closest  $(\sim A \wedge \sim C)$ -worlds, if these semifactuals are to be true. But this will not be the case for such conditionals because in such cases  $A$  is not sufficient to bring  $C$  about. Penczek, then, does not get the truth conditions correct for these irrelevant semifactuals (as he seems to accept 1997: 84 n12).<sup>xi</sup>

Of course, someone antecedently wedded to the Connection Hypothesis might reject that conditionals such as (3) are true. But the Connection Hypothesis is still to be rejected as it is incompatible with one of the most plausible principles of counterfactual logic, namely, Weakening the Consequent (Nute 1980: 7). This is because, although the Connection Hypothesis allows ‘if I were to push the button, the doorbell would ring’ to be true in suitable circumstances, it does not allow ‘if I were to push the button, the doorbell would exist’ to be true in those same circumstances, even though the doorbell’s ringing entails that the doorbell exists ( $\exists x$   $x$ =the doorbell).

Returning to Penczek’s semantics, we can illustrate the failure of Weakening the Consequent with the following model of worlds and propositions true at them (in this and the models that follow,  $W1$  is the world of evaluation) –  $W1: (A \wedge B \wedge C)$ ;  $W2: (\sim A \wedge \sim B \wedge C)$ ;  $W3: (\sim A \wedge \sim B$

$\wedge \sim C$ ); W4:  $(A \wedge \sim B \wedge \sim C)$ .  $A > B$  is true in virtue of the truth of its components, so the truth of  $A \Box \rightarrow B$  turns on the truth of  $(\sim A \wedge \sim B) > (A > B)$ . This is true in our model assuming that W2 is the closest  $(\sim A \wedge \sim B)$ -world to W1 and W1 is the closest A-world to W2. And B entails C, since all the B-worlds are C-worlds. But  $A \Box \rightarrow C$  need not be true in this model, since  $(\sim A \wedge \sim C) > (A > C)$  is false when W4 is amongst the closest A-worlds to W3. As a result a result Weakening the Consequent is invalid.

Penczek's proposal also invalidates Agglomeration as the following countermodel shows. W1:  $(A \wedge B \wedge C)$ ; W2:  $(\sim A \wedge \sim B \wedge C)$ ; W3:  $(\sim A \wedge B \wedge \sim C)$ ; W4:  $(A \wedge \sim B \wedge C)$ .  $A \Box \rightarrow B$  is true assuming W1 is the closest A-world to W2. Similarly,  $A \Box \rightarrow C$  is true since all the A-worlds are C-worlds. But  $A \Box \rightarrow (B \wedge C)$  is false if W3 is amongst the closest:  $(\sim A \wedge \sim(B \wedge C))$ -worlds to W1, and W4 is amongst the closest A-worlds to W3.

We should note that the above countermodels respect the following constraints that Lewis (1971) imposes on the closeness ordering:

Constraint 1: If the closest A-worlds are all B-worlds, and the closest B-worlds are A-worlds, then the closest A-worlds are the closest B-worlds.

Constraint 2: Either all the closest  $(A \vee B)$ -worlds are A-worlds, or all the closest  $(A \vee B)$ -worlds are B-worlds or the set of the closest  $(A \vee B)$ -worlds is the union of the set of the closest A-worlds and the closest B-worlds.

But perhaps the validity of Weakening the Consequent and Agglomeration could be restored by imposing further constraints on the closeness ordering. The following constraint validates the principles above

Constraint 3: For any  $A$  and  $C$  which are both true at  $w$ , if  $v$  is amongst the closest  $(\sim A \wedge \sim C)$ -worlds to  $w$ , then  $w$  is *the* closest  $(A \wedge C)$ -world to  $v$ .

Unfortunately, however, Constraint 3 also validates Conjunction Conditionalization as the world of evaluation is the only relevant  $A$ -world. The weaker

Constraint 4: For any  $A$  and  $C$  which is true at  $w$ , if  $v$  is amongst the closest  $(\sim A \wedge \sim C)$ -worlds to  $w$ , then  $w$  is *amongst* the closest  $(A \wedge C)$ -world to  $v$

does not validate Conjunction Conditionalization, but neither does it validate Weakening the Consequent and Agglomeration which can be seen by checking the countermodels above.

Of course, that Constraints 3 and 4 are of no help does not show that there is no constraint which delivers the appropriate logic. But if Penczek holds on to the Connection Hypothesis, then not only will his account give the incorrect verdict on irrelevant semifactuals, Weakening the Consequent will also be invalidated as we have seen.

In the light of the difficulties above, Penczek's account it is to be rejected, in its current form at least. (For the same reason we should also reject Bigelow's (1976) similar account:  $A \Box \rightarrow C$  is

true at  $w$  iff  $A > C$  is true throughout some sphere of worlds around  $w$  which includes a  $(\sim A \wedge \sim C)$ -world). As we noted, Penczek (1997: 84 n12) was alive to the problem of semifactuals, and although he doesn't explicitly address it, he notes that perhaps his truth condition could be modified to allow for irrelevant semifactuals such as (3) to be true. It is to such a modification that we now turn.

## 5. McGlynn

McGlynn (2012) seeks to modify Penczek's semantics in order to avoid the counterexamples from semifactuals discussed above. Cogburn and Roland (2013: 264 n11) "suspect that such a modification would be ad hoc", but there is an obvious, principled amendment to Penczek's semantics which allows for the truth of irrelevant semifactuals without Conjunction Conditionalization.

When  $A$  and  $C$  are false, Penczek's truth condition agrees with the standard account. But when at least one of  $A$  or  $C$  are true, we have to go to the closest  $(\sim A \wedge \sim C)$ -worlds, and assess whether or not  $A > C$  is true there, as well being true in the world of evaluation. As we saw above, this led to problems with irrelevant semifactuals - in such cases the truth of the consequent *is* relevant to the truth of the conditional, and so we should not be forced to consider  $\sim C$ -worlds. The obvious solution is to deviate from the standard account not when 'A or C' is true, but only when  $A$  is true. This is approach taken by McGlynn and he amends Penczek's account above, replacing (c) with

d.  $\sim A > (A > C)$ .

McGlynn's account, like Penczek's, invalidates Conjunction Conditionalization by forcing us to consider  $\sim A$ -worlds when  $A$  is true. But unlike Penczek, McGlynn allows for the truth of irrelevant semifactuals such as (3), since, given its independence from my betting behaviour, the result of the coin toss is held fixed at all the relevant worlds. Despite this result, and its intuitive appeal, McGlynn's account is subject to a number of difficulties.<sup>xii</sup>

First, McGlynn's treatment of conditionals with necessary antecedents is inadequate by the lights of those who reject Conjunction Conditionalization, since although it invalidates Conjunction Conditionalization, it does validate

Conjunction Conditionalization\*:  $(\Box A \wedge C) \supset (A \Box \rightarrow C)$ .

On McGlynn's semantics, when  $A$  and  $C$  are true,  $A \Box \rightarrow C$  is true, iff  $\sim A > (A > C)$ . But  $\sim A > (A > C)$  is vacuously true when  $A$  is necessary.

Although, Conjunction Conditionalization\* is weaker than Conjunction Conditionalization, it is not acceptable to those who reject Conjunction Conditionalization. The reason many reject Conjunction Conditionalization is that counterfactuals like (1) are prima facie false. But we have the same phenomena with counterfactuals with necessary antecedents and true consequents. For example, in McDermott's scenario above, (15) sounds at least as bad as (1):

(15) If none or more heads had come up, I would have won.

But if Conjunction Conditionalization\* is true, then (15) follows. It is clear that our parallel intuitive verdicts on (1) and (15) do not underwrite the asymmetric treatment McGlynn's account mandates.<sup>xiii</sup>

To avoid this problem, McGlynn would have to provide non-trivial truth conditions for  $A > C$  when A is necessary, which he could then employ in (d). The problem of counterfactuals with true antecedents has been replaced, then, by the problem of providing non-trivial truth conditions for counterfactuals with impossible antecedents. But McGlynn offers no account of these, and is it unclear how an account which suits McGlynn's purposes would go. McGlynn's account, then, is radically incomplete at best. Worse is to come, however.

As well as necessary antecedents, complex antecedents present a host of problems for McGlynn's account. The models below show that McGlynn's account leads to failures of SDA\*, Limited Transitivity, and Disjunction\*.

SDA\* - W1:  $(A \wedge \sim B \wedge C)$ ; W2:  $(\sim A \wedge B \wedge \sim C)$ ; W3:  $(A \wedge \sim B \wedge \sim C)$ ; W4:  $(\sim A \wedge \sim B)$ .  $A \Box \rightarrow C$  is false when W2 is the closest  $\sim A$ -world to W1, and W3 is amongst the closest A-worlds to W2.  $B \Box \rightarrow C$  is false when W2 is amongst the closest B-worlds to W1. And yet  $(A \vee B) \Box \rightarrow C$  is true when W1 is the closest  $(A \vee B)$ -world to W4.

Limited Transitivity - W1:  $(A \wedge B \wedge C)$ ; W2:  $\sim A$ ; W3:  $(A \wedge B \wedge \sim C)$ ; W4:  $(A \wedge \sim B)$ . If W1 and W3 are the closest A-worlds to W2, then  $A \Box \rightarrow B$ , but not  $A \Box \rightarrow C$ . If we then add that W4 is the

closest  $\sim(A \wedge B)$ -world to W1 and W1 the closest  $(A \wedge B)$ -world to W4,  $(A \wedge B) \Box \rightarrow C$  is also true.

Disjunction\* - W1:  $(A \wedge B \wedge C)$ ; W2:  $(\sim A \wedge B \wedge C)$ ; W3:  $(A \wedge \sim B \wedge C)$ ; W4:  $(\sim A \wedge \sim B)$ ; W5:  $(\sim A \wedge B \wedge \sim C)$ . If W2 and W3 are the closest  $\sim A/\sim B$ -worlds to W1 respectively, and W1 the closest  $(A \wedge B)/A/B$ -worlds to W2 and W3, then  $A \Box \rightarrow C$ ,  $B \Box \rightarrow C$ , and  $(A \wedge B) \Box \rightarrow C$ . But if the closest  $\sim(A \vee B)$ -world to W1 is W4, and amongst the closest  $(A \vee B)$ -worlds to it are W5, then  $(A \vee B) \Box \rightarrow C$  is false.<sup>xiv</sup>

So here we see that McGlynn's semantics invalidates a range of compelling principles of counterfactual logic. As with Penczek's account, appealing to Constraints 1 and 2 does not help. The validity of the principles can be restored on McGlynn's semantics if it is supplemented with.

Constraint 5: For any A which is true at w, if v is amongst the closest  $\sim A$ -worlds to w, then w is *the* closest A-world to v.

Unfortunately, however, Constraint 5 also validates Conjunction Conditionalization as the world of evaluation is the only relevant A-world.<sup>xv</sup> The weaker

Constraint 6: For any A which is true at w, if v is amongst the closest  $\sim A$ -worlds to w, then w is *amongst* the closest A-world to v

does not validate Conjunction Conditionalization, but neither does it validate the theorems of VW which can be seen by checking the countermodels above.

Of course, this does not show that there is no constraint which delivers the appropriate logic. But given McGlynn's other commitments we have good reason to think that no such constraint will be forthcoming. Consider the countermodel to Disjunction\* above. Let us add that in W1, C is overdetermined by two independent indeterministic causes, A and B. Given this set up the closest  $\sim A$ -world to W1 will be W2. This is because whether or not A obtains is irrelevant to the causal chain between B and C, and so we hold the result of this causal chain fixed when considering  $\sim A$ . By parallel reasoning W3 is the closest  $\sim B$ -world to W1. As C was caused by both A and B, and absent any pre-empted causes, the closest  $\sim(A \vee B)$ -worlds to W1 will be ones like W4, where  $\sim C$ . Now since A and B are *indeterministic* causes of C, the closest  $(A \vee B)$ -worlds to W4 will include worlds like W5 where at least one of the causes is present but the effect is not. So given the truth of irrelevant semifactuals even in indeterministic contexts, McGlynn's semantics cannot be supplemented to validate Disjunction\*.

It might be replied on McGlynn's behalf that he manages to hold on to the above principles when the antecedents are false and that this is good enough. But principles like Disjunction\* seem compelling regardless of the truth value of the antecedents involved. Similarly, the good instances of Transitivity, to be explained by the validity of Limited Transitivity, look good even when the antecedents are true. But McGlynn can't help himself to this explanation since he invalidates Limited Transitivity in such cases. The possible world semanticist seeking to invalidate Conjunction Conditionalization must, then, consider alternative proposals.

## 6. Cogburn and Roland

Cogburn and Roland (2013: 257), motivated by problems with Sosa's counterfactual account of safe belief, provide an alternative semantics for counterfactuals which avoids Conjunction Conditionalization.<sup>xvi</sup> Their proposal is that:

$A \Box \rightarrow C$  is non-vacuously true in a context  $C$  and a world  $w$  iff most of the  $(C,w)$ -relevantly similar  $A$ -worlds are  $C$ -worlds.<sup>xvii</sup>

Cogburn and Roland's idea is that within a context, there is some fixed class of worlds that are relevant to the truth of counterfactuals in that context. If most of the  $A$ -worlds within this set are  $C$ -worlds, then  $A \Box \rightarrow C$ . Their proposal is akin to a contextually-sensitive strict conditional account of counterfactuals, such as von Fintel's (2001), except that for Cogburn and Roland the truth of  $A \Box \rightarrow C$  requires only that *most*  $A$ -worlds are  $C$ -worlds, not that all of them are.<sup>xviii</sup> An alternative proposal along Cogburn and Roland's lines is to amend Lewis's variably strict approach:  $A \Box \rightarrow C$  is non-vacuously true iff most of the closest  $A$ -worlds are  $C$ -worlds. Nothing I say below turns on whether we opt for Cogburn and Roland's official account or the variant just sketched.

Although it avoids Conjunction Conditionalization, Cogburn and Roland's proposal is manifestly flawed. To start with, Cogburn and Roland's specific proposal has two obvious problems. First, rather than 'most', Cogburn and Roland would do better to opt for 'at least  $n\%$ ', for some large  $n$  (cf. Bennett, 2003: §98). This is because, if I do not toss a biased coin which comes up heads

51% of the time, we do not want to endorse ‘if I had tossed the coin, it would have come up heads’ even though most of the relevant worlds in which it is tossed, are ones in which it lands heads.

Second, Cogburn and Roland’s proposal invalidates Modus Ponens (cf. Bennett, 2003: 250). That  $n\%$  of  $(C,w)$ -similar A-worlds are C-worlds does not preclude the actual world being an  $(A \wedge \sim C)$ -world. Although Modus Ponens for counterfactuals is occasionally rejected, most theorists who reject Conjunction Conditionalization wish to validate it. Moreover, Cogburn and Roland’s motivation for rejecting Conjunction Conditionalization does not extend to Modus Ponens.

It is easy enough to amend Cogburn and Roland’s proposal in the light of the above problems, however, and some of their remarks suggest the way that they would go:

A  $\Box \rightarrow C$  is non-vacuously true in a context C and a world w iff at least  $n\%$  of  $(C,w)$ -relevantly similar A-worlds are C-worlds and all very close  $(C,w)$ -relevantly similar A-worlds are C-worlds.

If we add to this truth condition that the actual world is always one of the very close  $(C,w)$ -relevantly similar A-worlds for any true A, Modus Ponens is restored.<sup>xix</sup>

But even this tidied-up version of Cogburn and Roland’s proposal is inadequate.<sup>xx</sup> The logic of ‘at least  $n\%$ ’ is very different from the logic of ‘all’, so we should expect Cogburn and Roland’s semantics to deliver a logic different to the standard Stalnaker-Lewis logic. And this is indeed

the case. The most obvious difference stems from the invalidity of: at least  $n\%$  of A's are B; at least  $n\%$  of A's are C; therefore, at least  $n\%$  of A's are B and C. This feature of the logic of 'at least  $n\%$ ' means that Cogburn and Roland's proposal invalidates Agglomeration (cf. Hawthorne, 2003: 397-398). This alone means that Cogburn and Roland's account represents a significant departure from the standard account. But it is worth noting that their account is revisionary in further ways.

First, Cogburn and Roland's proposal invalidates Limited Transitivity: that  $n\%$  of A-worlds are B-worlds, and  $n\%$  of (A  $\wedge$  B)-worlds are C-worlds, only entails that  $n\%^2$  of A-worlds are C-worlds.

Second, their proposal invalidates Disjunction\*: suppose that there are  $n$ -worlds where (A  $\wedge$  B  $\wedge$  C), one world where (A  $\wedge$   $\sim$ B  $\wedge$   $\sim$ C), and one world where ( $\sim$ A  $\wedge$  B  $\wedge$   $\sim$ C). This means that  $n/(n+1)\%$  of A-worlds are C-worlds,  $n/(n+1)\%$  of B-worlds are C-worlds, and 100% of (A  $\wedge$  B)-worlds are C-worlds, but only  $n/(n+2)\%$  of (A  $\vee$  B)-worlds are C-worlds. So, if it is required that  $n/(n+1)\%$  of antecedent-worlds are consequent worlds for a counterfactual to be true, then A  $\Box \rightarrow$  C, B  $\Box \rightarrow$  C, and (A  $\wedge$  B)  $\Box \rightarrow$  C are all true, whereas (A  $\vee$  B)  $\Box \rightarrow$  C is not, and so Disjunction\* is invalid.

Finally, Cogburn and Roland have to reject SDA\* too as the following model shows: let there be  $m$  worlds which are (A  $\wedge$   $\sim$ B  $\wedge$  C)-worlds,  $m$  worlds which are ( $\sim$ A  $\wedge$  B  $\wedge$  C)-worlds, and  $n$  worlds which are (A  $\wedge$  B  $\wedge$   $\sim$ C)-worlds. Then we have it that there are  $2m/(n+2m)$  ((A  $\vee$  B)  $\wedge$  C)-worlds, but only  $m/(n+m)$  (A  $\wedge$  C)-worlds and  $m/(n+m)$  (B  $\wedge$  C)-worlds. But  $2m/(n+2m)$  is

always greater than  $m/(n+m)$ , so if we set the threshold for truth at  $2m/(n+2m)\%$ , then  $(A \vee B) \Box \rightarrow C$  will be true, but neither of  $A \Box \rightarrow C$  or  $B \Box \rightarrow C$  will be, thus invalidating SDA\*.

So, like the accounts above, Cogburn and Roland invalidate Limited Transitivity, SDA\*, and Disjunction\*, and in addition they also invalidate Agglomeration. This seems like a large cost just to avoid Conjunction Conditionalization.

## 7. Gundersen

Gundersen (2004) rejects Conjunction Conditionalization for reasons along the lines of the Connection Hypothesis. What he thinks is important for  $A \Box \rightarrow C$  is that all 'normal' A-worlds be C-worlds. And the type of normality or connection that he thinks is important is probability raising. So, for Gundersen,  $A \Box \rightarrow C$  is true iff both of the following are satisfied:

(16)  $P(C/A)$  is greater than  $P(\sim C/A)$

(17)  $P(C/A)$  is greater than  $P(C/\sim A)$ <sup>xxi</sup>

As we noted above (§4), those who endorse the Connection Hypothesis face a general challenge, namely that irrelevant semifactuals seem to provide a counterexample, and relatedly, that the Connection Hypothesis violates Weakening the Consequent. Gundersen's response is to distinguish between genuine counterfactuals and subjunctives with true antecedents on the one hand, and semifactuals on the other. For Gundersen, the former types of conditional assert a connection between the antecedent and the consequent, and are to be treated by his clause for  $A \Box \rightarrow C$  above. On the other hand, Gundersen claims that semifactuals assert the lack of the

opposite connection. That is, a semifactual ‘if A were the case, C would have been the case’ is treated by Gundersen not as  $A \Box \rightarrow C$ , but rather as  $\sim(A \Box \rightarrow \sim C)$ . As a result Gundersen need not reject the truth of irrelevant semifactuals, nor does his account invalidate Weakening the Consequent *on the grounds that Penczek’s account does*.<sup>xxii</sup> Nevertheless, Gundersen’s account is not without its problems.

Given that  $P(C/A)=n$  says that the proportion of the relevant A-cases that are C-cases is n, the logic that Gundersen’s semantics validates shares certain features with the logic of Cogburn and Roland’s account. In particular, neither Modus Ponens nor Agglomeration is validated. Gundersen is aware of the former and takes this to be a virtue of his account, since w can be abnormal, for instance by being an  $(A \wedge \sim C)$ -world when A almost invariably brings about C. For Gundersen this latter fact makes  $A \Box \rightarrow C$  true at w. Whatever the merits of dropping Modus Ponens, invalidating Agglomeration seems like a large cost.

Leaving to one side Modus Ponens and Agglomeration, Gundersen’s account faces further problems. First, Gundersen has failed to provide an account of subjunctives with necessary antecedents, since  $P(C/\sim A)$  is undefined when A is necessary.

Second, Gundersen’s semantics has further unpalatable consequences for the logic of conditionals as it invalidates Disjunction, SDA\*, and Limited Transitivity in essentially the same way as Cogburn and Roland’s does since  $P(C/A)$  is greater than  $P(\sim C/A)$  iff most of the A-worlds are C worlds. But in addition, Gundersen’s account also invalidates Weakening the Consequent. This is because when B entails C, the following claims are consistent:  $P(C/A)$  is

greater than  $P(\sim C/A)$ ;  $P(B/A)$  is greater than  $P(\sim B/A)$ ;  $P(B/A)$  is greater than  $P(B/\sim A)$ ; and yet  $P(C/A)$  is not greater than  $P(C/\sim A)$ . This is because all of the  $\sim A$ -worlds could be  $C$ -worlds, but only a few of them be  $B$ -worlds.

Gundersen emphasises that he is proposing an amendment of the standard account (2004: 3) and (2004: 18) that there is “a considerable degree of agreement” between his theory and Lewis’s. Here I have highlighted the differences between the two theories and I take this to count against Gundersen.<sup>xxiii</sup>

## **8. Lessons**

We have seen that the above proposals for invalidating Conjunction Conditionalization have wider consequences for the logic of counterfactuals. Moreover, although they are within the spirit of the standard approach, the resulting logic is importantly different. There is, then, a general lesson that the above attempts to circumvent Conjunction Conditionalization remind us of. One cannot simply take a semantics,  $S$ , which validates a logic,  $L$ , a theorem of which is  $P$ , and amend  $S$  to some  $S^*$  which does not validate  $P$ , so that  $S^*$  validates the logic that results from subtracting  $P$  from  $L$ .

There is, in addition, a more localized lesson. Penczek’s and McGlynn’s accounts mishandle counterfactuals with complex antecedents because they, in effect, make the truth of a counterfactual an antecedent-relative matter. On the standard account, antecedents determine that we are concerned only with worlds where the antecedent holds true

but that is their only role in determining [the] selection [of worlds]. The rest of the job is done by some antecedent-independent conception of similarity or minimal difference (Stalnaker, 1987: 129-130).<sup>xxiv</sup>

On McGlynn's semantics, however, this is not the case: when an antecedent,  $A$ , is true, it plays a further role, partly determining which worlds are relevant to determining the conditional's truth. For example, when  $(A \wedge B)$ , the worlds relevant to assessing  $A \Box \rightarrow C$  can be disjoint from the worlds relevant to assessing  $B \Box \rightarrow C$ , even when both sets of worlds are  $(A \wedge B)$ -worlds. And similar remarks apply to Penczek's account. True antecedents, then, do not just determine that we are concerned only with antecedent worlds. Such antecedent-relative semantics, without further constraints, do not validate the logical principles above in virtue of their antecedent-relativity. Those who wish to endorse antecedent-relativity should, then, ensure that they constrain their semantics in ways which deliver the appropriate logic.<sup>xxv</sup>

Although my focus has been on counterfactuals, it is worth noting a general lesson here for epistemologists. Cogburn and Roland, Gundersen, and McGlynn, all motivate their semantics, in part, by the utility of their semantics for epistemology. But Cogburn and Roland's final version of safe belief (cf. Pritchard, 2009: 34) makes no mention of counterfactuals:

S's belief that  $P$  is safe iff in *most* relevant worlds where  $S$  believes  $P$  (SBP),  $P$ , and in *all* very close relevant worlds where SBP,  $P$ .

More generally, we can specify an account of safe belief without committing to any claim about the semantics of counterfactuals: if we think  $T$  gives the truth conditions of  $SBP \Box \rightarrow P$ , rather than saying  $SBP$  is safe iff  $SBP \Box \rightarrow P$ , we can say instead that  $SBP$  is safe iff  $T$ , cutting out the claim about the semantics of counterfactuals. In McGlynn's terms, we can say that  $SBP$  is safe iff  $\sim SBP > (SBP > P)$ . And what goes for safe belief goes for sensitive belief too. Formulating safety and sensitivity in terms of natural language counterfactuals adds nothing to a theory of knowledge, is unnecessarily committal, and so is hostage to the kind of issues we've been discussing above. This is particularly clear in the case of Cogburn and Roland, as their proposal simply grafts an implausible semantics for counterfactuals onto something like Pritchard's account of safe belief. Rather than amending the semantics of counterfactuals to make them suitable for epistemology, epistemologists should simply pick another tool for the job. Of course, whether safety and sensitivity are useful in epistemology is another question. But whether they are, or are not, does not turn on making substantial claims about natural language semantics.

## **9. Revisiting the Problem**

We have seen that a range of possible worlds semantics for counterfactuals which invalidate Conjunction Conditionalization have interesting consequences that many will take to show their inadequacy. Given the difficulties with invalidating Conjunction Conditionalization within a logically conservative possible world semantics, perhaps we should abandon thinking of counterfactuals in terms of possible worlds (cf. Fine 2012). Alternatively, we could re-examine our reasons for rejecting Conjunction Conditionalization in the first place.

We can recognize the validity of Conjunction Conditionalization, whilst at the same time accepting that not all counterfactuals with true components are assertable. Gundersen (2004: 3) claims that “counterfactuals are associated with a pragmatic implicature: one ought to assert them when their ... antecedent as a matter of fact is false. Or, at least one should only assert them when one, for whatever reasons, takes the antecedent to be false”.<sup>xxvi</sup> This isn’t quite right, however. As Edgington (2004: 23) notes, ‘I think he must have taken arsenic, since he has such-and-such symptoms, and these are just the symptoms he would have, if he had taken arsenic’ is perfectly felicitous. Rather, what I want to suggest is that counterfactuals are assertable only if they are true regardless of the truth value of the antecedent, so that McGlynn’s truth conditions are in fact assertability conditions.

If Conjunction Conditionalization is valid, then  $A \wedge C$  is strictly stronger than  $A \Box \rightarrow C$ . As a result, we should, by Grice’s maximum of quantity, *ceteris paribus* assert  $A \wedge C$ , if we know it, rather than the weaker  $A \Box \rightarrow C$ . Of course, some counterfactuals with true components are assertable even when the truth of  $A \wedge C$  is known, but given that  $A \wedge C$  is strictly stronger than  $A \Box \rightarrow C$ , there must be some reason to assert the weaker  $A \Box \rightarrow C$ . I suggest that if  $A \Box \rightarrow C$  is assertable when  $A \wedge C$  is known, it is because we are not relying on the truth of the antecedent for the truth of the counterfactual so that the conditional is true regardless of the truth of  $A$ . In particular, when  $A \Box \rightarrow C$  is assertable in such cases,  $\sim A > (A > C)$  is true as well. So an assertion of  $A \Box \rightarrow C$  conversationally implies or otherwise suggests  $\sim A > (A > C)$ . But when  $\sim A > (A > C)$  is false, as it is in the problematic instances of Conjunction Conditionalization,  $A \Box \rightarrow C$  is not assertable. On the other hand, if we do not know that  $A$ , then in the disputed instances of

Conjunction Conditionalization we do not know  $A \Box \rightarrow C$  either, since  $A \Box \rightarrow C$  is true in these cases only when  $A \wedge C$  is true. That is, in such cases  $\sim A > (A > C)$  is false, and since we cannot rule out  $\sim A$ , the falsity of  $(A > C)$  is an epistemic possibility. Therefore, on the standard account, where  $(A > C) \equiv (A \Box \rightarrow C)$ ,  $A \Box \rightarrow C$  will be unassertable given either Grice's maxim of quality or a knowledge norm of assertion. So whether or not we know  $A \wedge C$ ,  $A \Box \rightarrow C$  is unassertable in the problematic cases. So the advocate of Conjunction Conditionalization can appeal to the falsity of  $\sim A > (A > C)$  to explain the infelicity of some true-true counterfactuals just as McGlynn does, but without encountering the problems that his account generates.<sup>xxvii</sup>

Department of Philosophy

University of Southampton

## Appendix

Disjunction:  $((A \Box \rightarrow C) \wedge (B \Box \rightarrow C)) \supset ((A \vee B) \Box \rightarrow C)$

Disjunction\*:  $((A \Box \rightarrow C) \wedge (B \Box \rightarrow C) \wedge ((A \wedge B) \Box \rightarrow C)) \supset ((A \vee B) \Box \rightarrow C)$ .

Disjunction obviously entails Disjunction\*, so all that remains is to show the reverse entailment in a background logic of counterfactuals consisting of Agglomeration, Weakening the Consequent, Reflexivity, and Substitution of Logical Equivalents:

- |   |                          |
|---|--------------------------|
| 1. $(A \Box \rightarrow C) \wedge (B \Box \rightarrow C)$ | Assumption               |
| 2. $A \Box \rightarrow C$                                 | 1, $\wedge$ -Elimination |

3. $B \Box \rightarrow C$	1, $\wedge$ -Elimination
4. $A \Box \rightarrow (C \vee (B \wedge \sim A))$	2, Weakening the Consequent
5. $(B \wedge \sim A) \Box \rightarrow (B \wedge \sim A)$	Reflexivity
6. $(B \wedge \sim A) \Box \rightarrow (C \vee (B \wedge \sim A))$	5, Weakening
7. $(A \wedge B \wedge \sim A) \Box \rightarrow (A \wedge B \wedge \sim A)$	Reflexivity
8. $(A \wedge B \wedge \sim A) \Box \rightarrow (C \vee (B \wedge \sim A))$	7 Weakening
9. $(A \vee (B \wedge \sim A)) \Box \rightarrow (C \vee (B \wedge \sim A))$	4,6,8, Disjunction*
10. $(A \vee B) \Box \rightarrow (C \vee (B \wedge \sim A))$	9, Substitution of logical equivalents

By a symmetrical argument (switching A and B everywhere) we have

$$11. (A \vee B) \Box \rightarrow (C \vee (A \wedge \sim B))$$

so this gives us

12. $(A \vee B) \Box \rightarrow ((C \vee (A \wedge \sim B)) \wedge (C \vee (B \wedge \sim A)))$	10,11, Agglomeration
13. $(A \vee B) \Box \rightarrow C$	12, Weakening the Consequent
14. $((A \Box \rightarrow C) \wedge (B \Box \rightarrow C)) \supset ((A \vee B) \Box \rightarrow C)$	1, 13, Conditional Proof. <sup>xxviii</sup>

## References

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

Ahmed, A. (2011) "Walters on Conjunction Conditionalization". *Proceedings of the Aristotelian Society* CXI: 113-20.

Bennett, J. (1974) "Counterfactuals and Possible Worlds". *Canadian Journal of Philosophy* 4: 381-402.

Bennett, J. (2003) *A Philosophical Guide to Conditionals*. Oxford: OUP.

Bigelow, J. (1976) "If-then meets the possible worlds". *Philosophia* 6: 215-236.

Burgess, J. (1981) "Quick Completeness Proofs for Some Logics of Conditionals". *Notre Dame Journal of Formal Logic* 22: 76-84.

Butcher, D. 1983: An Incompatible Pair of Subjunctive Conditional Modal Axioms. *Philosophical Studies*, 44: 71-110.

Cogburn, J. and Roland, J.W. (2013) "Safety and the True-True Problem". *Pacific Philosophical Quarterly* 94: 246-267.

Dowe, P. and Noordhof, P. (eds) (2004) *Cause and Chance: Causation in an Indeterministic World*. London: Routledge.

Edgington, D. (2004) "Counterfactuals and the Benefit of Hindsight". In Dowe, P. and Noordhof, P. (eds) 2004, pp.12-27.

Fine, K. (1975) "Critical Notice of David Lewis's Counterfactuals". *Mind* 84: 451-458.  
Reprinted in his *Modality and Tense* (2005) Oxford: OUP, pp. 357-365

Fine, K. (2012) "Counterfactuals without Possible Worlds". *Journal of Philosophy* CIX: 221-246.

von Fintel, K. (2001) "Counterfactuals in a Dynamic Context". In *Ken Hale: A Life in Language*. Edited by Michael Kenstowicz. Cambridge, MA: MIT Press, pp. 123-92.

Gärdenfors, P. (1978) "Conditionals and changes of belief". *Acta Philosophica Fennica* 30: 381-404.

Gauker, C. (2005) *Conditionals in Context*. Cambridge MA: MIT Press.

Gundersen, L. (2004) "Outline of a new semantics for counterfactuals". *Pacific Philosophical Quarterly* 85: 1-20.

Hawthorne, J. (2005) "Chance and counterfactuals". *Philosophy and Phenomenological Research* 70: 396-405.

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

Lewis, D. (1971) “Completeness and decidability of three logics of counterfactual conditionals”. *Theoria* 37: 74-85.

Lewis, D. (1979) “Counterfactual Dependence and Time’s Arrow”. *Nous* 13: 455-76. Reprinted with postscripts in his *Philosophical Papers: Volume II*. (1986). Oxford: OUP.

Lewis, D. (1986) *Counterfactuals*, revised edn. Oxford: Basil Blackwell.

Lowe, E.J. (1995) “The Truth About Counterfactuals”. *The Philosophical Quarterly* 45: 41-59.

McDermott, M. (2007) “True Antecedents”. *Acta Analytica* 22: 333-335.

McGlynn, A. (2012) “The Problem of True-True Counterfactuals”. *Analysis* 72: 276-285.

McKay, T. and van Inwagen, P. (1977) “Counterfactuals with disjunctive antecedents”. *Philosophical Studies* 31: 353-356.

Mizrahi, M. (2013) “Why Hypothetical Syllogism is Invalid for Indicative Conditionals”. *Thought* 2: 40-43.

Noordhof, P. (2004) “Prospects for a counterfactual theory of causation”. In Dowe and Noordhof (eds) 2004, pp. 188–201.

Nute, D. (1980) *Topics in Conditional Logic*. Boston, MA: Reidel.

Penczek, A. (1997) “Counterfactuals with true components”. *Erkenntnis* 46: 79-85.

Phillips, I. (2007) “Morgenbesser cases and closet determinism”. *Analysis* 67: 42-49.

Pritchard, D. (2009) “Safety-Based Epistemology, Whither Now?” *Journal of Philosophical Research* 34: 33–45.

Pollock, J. (1976) *Subjunctive Reasoning*. Boston, MA: Reidel.

Pruss, A. (2007) “Conjunctions, Disjunctions and Lewisian Semantics for Counterfactuals”. *Synthese* 96: 33-52.

Schaffer, J. (2004) “Counterfactuals, causal dependence and conceptual circularity”. *Analysis* 54: 299-309.

Stalnaker, R. (1987) *Inquiry*. Cambridge, MA: MIT Press.

Tichý, P. (1978) “A New Theory of Subjunctive Conditionals”. *Synthese* 37: 433-457.

Tooley, M. (2002) “Backward causation and the Stalnaker–Lewis approach to counterfactuals”. *Analysis* 62: 191–197.

Walters, L. (2009) “Morgenbesser's Coin and Counterfactuals with True Components”. *Proceedings of the Aristotelian Society* CIX: 365-379.

Walters, L. (2011) “Reply to Ahmed”. *Proceedings of the Aristotelian Society* CXI: 123-133.

Walters, L. (2014a) “Against Hypothetical Syllogism”. *Journal of Philosophical Logic* 43: 979-997.

Walters, L. (2014b) “Conditionals, Modals, and Hypothetical Syllogism”. *Thought* 3: 90-97.

Walters, L. and Williams, J.R.G. (2013) “An Argument for Conjunction Conditionalization”. *The Review of Symbolic Logic* 6: 573-588.

---

<sup>i</sup> Of course, that similarity is here a technical notion may not have been clear at the time Bennett was writing.

<sup>ii</sup> To be fair to McGlynn, he seems to have multiple reasons for rejecting Lewis's suggestion since as well as citing Bennett (1974), he also cites McDermott (2007) who rejects Lewis's suggested retreat to Weak Centring for the reason discussed below.

<sup>iii</sup> Cogburn and Roland (2013: 246) compound the error, claiming that “since by any measure of similarity no world is more similar to the actual world than the actual world itself, it follows that the truth value of a counterfactual conditional with a true antecedent will be the truth value of the conditional's consequent”. But this is a non-sequitur: what they have characterised here is Weak Centring, and Conjunction Conditionalization does not follow from that, even given (i).

<sup>iv</sup> This is the position of Bennett (2003). Bennett provides his own variant of the standard possible world semantics, which I discuss elsewhere (Walters: 2009). See also n15 below.

<sup>v</sup> Disjunction\* is equivalent to Disjunction assuming Agglomeration, Weakening the Consequent, that logical equivalents can be substituted in the antecedents of counterfactuals *salva veritate*, and Reflexivity:  $A > A$  (see appendix). But as we are considering semantic proposals which are logically revisionary we cannot assume the equivalence of Disjunction\* and Disjunction here. The following is a possible worlds countermodel to Disjunction but not Disjunction\*: the closest A-worlds are C-worlds, the closest B-worlds are C-worlds, and the closest  $(A \wedge B)$ -

---

worlds are a subset of the closest  $(A \vee B)$ -worlds all of which are  $\sim C$ -worlds. In the standard semantics the closeness relation is constrained in such a way as to rule out such countermodels (see Constraints 1 and 2).

<sup>vi</sup> I lack the space to discuss them here, although see Walters (2011) for a reply to Ahmed.

<sup>vii</sup> Assuming a weak background logic of counterfactuals and the principle  $(X \Box \rightarrow \perp) \supset \Box \sim X$ .

<sup>viii</sup> Fine (2012) argues that given some plausible non-logical assumptions, one cannot maintain Limited Transitivity, substitution of logical equivalents in the antecedents of counterfactuals, Disjunction, an infinitary version of Agglomeration, and the fact that counterfactuals whose antecedents entail their consequents are true. One could take Fine's argument as a reductio of Limited Transitivity, but obviously this is not the only option.

<sup>ix</sup> Gundersen (2004: 18) notes this. Cogburn and Roland (2013) explicitly design a semantics for which contraposition fails, and as a result it also invalidates Transitivity. Penczek's (1997) and McGlynn's (2012) semantics agree with Lewis's when the antecedent is false, and thus generate counterexamples to Transitivity when considering conditionals with false antecedents.

<sup>x</sup> It is not universally conceded that such examples show that Transitivity is invalid. I (Walters 2014a) present a different case against Transitivity, as does Mizrahi (2014), but Mizrahi's case is to be rejected (see Walters (2014b)).

<sup>xi</sup> Bennett (2003) and Fine (2012) now endorse the truth of irrelevant semifactuals and so reject the Connection Hypothesis.

<sup>xii</sup> Penczek's and Bigelow's accounts also face the difficulties for McGlynn's account highlighted below, and for the same type of reason. I'll not pause to demonstrate this fact though.

<sup>xiii</sup> The problem ramifies for Penczek, since the same problem arises when the antecedent and consequent are subcontraries:  $A \Box \rightarrow (\sim A \vee C)$  comes out as vacuously true when  $A \wedge C$  since for Penczek it is true iff  $(\sim A \wedge \sim(\sim A \vee C)) > (\sim A \vee C)$ . But as the antecedent of this latter conditional is impossible, this Lewisian conditional is vacuously true.

<sup>xiv</sup> McGlynn's semantics also invalidates the widely accepted  $((A \Box \rightarrow B) \wedge \Box(B \supset C)) \supset (A \Box \rightarrow C)$ , if the accessibility relation is not transitive as the following model shows - W1:  $(A \wedge B \wedge C)$ ; W2:  $\sim A$ ; W3:  $(A \wedge B \wedge \sim C)$ . If W3 is amongst the closest A-worlds to W2 then although  $A \Box \rightarrow B$  is true,  $A \Box \rightarrow C$  is not. And this is consistent with  $\Box(B \supset C)$  if W3 is inaccessible from W1, which it can be unless the accessibility relation is transitive.

<sup>xv</sup> Bennett (2003: 241), who rejects Conjunction Conditionalization, in effect endorses McGlynn's semantics. Bennett claims that if A and C are both true at w then  $A \Box \rightarrow C$  is true at w just in case it is true at the closest  $\sim A$ -worlds to w. Bennett also endorses Constraint 5: "Given that w is  $\alpha$ 's closest  $\sim A$ -world, presumably  $\alpha$  is w's closest A-world". But as we have just shown, with Constraint 5 in place McGlynn's semantics validates Conjunction Conditionalization. In Bennett's terms, if w is the closest  $\sim A$ -world to  $\alpha$ , then  $A \Box \rightarrow C$  is true at w, since the closest A-world to w is  $\alpha$  which is an  $(A \wedge C)$ -world. Given that  $A \Box \rightarrow C$  is true at w, the closest  $\sim A$ -world to  $\alpha$ , it is also true at  $\alpha$ . Generalizing, we have the validity of Conjunction Conditionalization. Thanks here to Charlie Temperley for reminding me of Bennett's discussion.

<sup>xvi</sup> McGlynn (2012: 284) and Gundersen (2004:11) also partly motivate their accounts in terms of the role counterfactuals are given in epistemology.

<sup>xvii</sup> Vacuous truth is not relevant to my discussion of Cogburn and Roland.

<sup>xviii</sup> Why do I not consider such context-sensitive strict conditional accounts of counterfactuals even though they invalidate Conjunction Conditionalization? First, these accounts do not invalidate VW and so an extended discussion of them would be out of place here. Second, although such accounts invalidate Conjunction Conditionalization, they still license instances which are problematic by the lights of those who argue against Conjunction Conditionalization. For instance, in a context in which no other counterfactuals have been asserted, von Fintel's account makes any counterfactual with true components true. Finally, I have argued against such accounts on independent grounds elsewhere (Walters: 2014a).

<sup>xix</sup> The final version of safe belief that Cogburn and Roland (2013: 260) adopt naturally suggests this semantics, although they don't propose it themselves.

<sup>xx</sup> The following criticisms all apply to the original proposal too.

<sup>xxi</sup> Although Gundersen puts things in terms of normal worlds, his semantics need not appeal to possible worlds at all as (16) and (17) do all the work.

<sup>xxii</sup> One might worry about Gundersen's proposal to distinguish between these conditionals on the following grounds. First, why group true-true counterfactuals with false-false counterfactuals in stating a connection, rather than with semifactuals in denying the opposite connection? Second, whilst people have thought that the truth of the

---

antecedent should not count in favour of a counterfactual being true, they have not traditionally thought that it should count against it being true. But this is what Gundersen effectively does, since he thinks that it is easier for ‘if A were the case, C would be the case’ to be true when it is a semifactual than when it is a true-true subjunctive: the semifactual only denies the lack of a connection between A and  $\sim$ C, whereas the latter asserts a connection between A and C. Finally, surely one can be certain that if he had scratched his nose, the coin would (still) have landed heads, regardless of the truth value of the antecedent? Gundersen’s treatment does not allow for that.

<sup>xxiii</sup> Although Gundersen’s account is flawed, those, like Gundersen, who wish to avoid both Conjunction Conditionalization and Modus Ponens within a possible world semantics, could simply adopt the standard semantics but reject both Strong and Weak Centering.

<sup>xxiv</sup> It is Constraints 1 and 2 that ensure this is the case.

<sup>xxv</sup> Ahmed (2011), who rejects Substitution and VLAS, explicitly endorses an antecedent-relative closeness ordering. I show (Walters 2011: §3), however, that Ahmed’s reasoning against these principles is invalid given his preferred semantics. The obvious constraint on the closeness ordering which validates Ahmed’s reasoning also validates Substitution and VLAS. So with or without this constraint, Ahmed’s argument against these principles is unsound.

<sup>xxvi</sup> Compare Lewis (1986: 3) “the counterfactual constructions of English do carry some sort of presupposition that the antecedent is false. It is some sort of mistake to use them unless the speaker does take the antecedent to be false”.

<sup>xxvii</sup> Thanks to three referees for very helpful comments which resulted in a much improved paper.

<sup>xxviii</sup> This proof is in very large part due to a slightly more committal proof of Andrew Bacon’s. I am grateful to Andrew for showing me the connection between Disjunction and Disjunction\*