

# Counterfactuals, Counteractuals, and Free Choice

Fabio Lampert<sup>1</sup> and Pedro Merluzzi<sup>2</sup>

<sup>1</sup>Department of Logic and Philosophy of Science - University of California, Irvine

<sup>2</sup>Center for Logic, Epistemology and the History of Science - University of Campinas

Preprint of February 2020. Forthcoming in *Philosophical Studies*.

## Abstract

In a recent paper, Pruss (2013) proves the validity of the rule BETA-2 relative to Lewis's semantics for counterfactuals, which is a significant step forward in the debate about the consequence argument. Yet, we believe there remain intuitive counter-examples to BETA-2 formulated with the actuality operator and rigidified descriptions. We offer a novel and two-dimensional formulation of the Lewisian semantics for counterfactuals and prove the validity of a new transfer rule according to which a new version of the consequence argument can be formulated. This new transfer rule is immune to the counter-examples involving the actuality operator and rigidified descriptions. However, we show that counter-examples to this new rule can also be generated, demanding that the Lewisian semantics be generalized for higher dimensions where counter-examples can always be generated.

## 1 Introduction

Central to van Inwagen's (1983) Consequence Argument for the incompatibility of determinism and free will is the much discussed rule BETA, a transfer principle asserting that if  $\varphi$  and there is no choice about whether  $\varphi$ ,  $\varphi$  implies  $\psi$  and there is no choice about whether  $\varphi$  implies  $\psi$ , then  $\psi$  and there is no choice about whether  $\psi$ , too. A decisive counter-example to BETA was given by McKay and Johnson (1996), but Widerker (1987) and Finch and Warfield (1998) suggested a stronger principle deemed BETA-2 which is not subject to

---

McKay and Johnson’s counter-example, and that allows an even simpler formulation of the consequence argument. Even so, the fact that BETA-2 was not immune to similar counter-examples was not enough to show that there was no possible counter-example to it. Yet, Pruss (2013) offered a proof of the validity of a formal rendering of BETA-2 against the background of the Lewisian semantics for counterfactuals and the modal logic **T**. Thus it might seem that the discussion surrounding the plausibility of BETA-2 is closed once and for all: the rule is valid with respect to the class of (reflexive and) relational models with the usual Lewisian ordering of possible worlds for a propositional modal language containing the counterfactual connective.

In this paper, however, we claim that even though there is nothing wrong with Pruss’s proof, BETA-2 is intuitively invalid: there are counter-examples to it that cannot be seen, as it were, by the formal apparatus, until the modal language is extended with an actuality operator—although we also argue that similar counter-examples can be generated with rigidified descriptions. In the interest of preserving the consequence argument though, we offer a novel and two-dimensional formulation of the Lewisian semantics for counterfactuals and prove the validity of a new transfer rule according to which a new version of the consequence argument can be formulated. This new transfer rule is immune to the counter-examples. However, we argue that similar counter-examples can still be generated, requiring a generalization of the Lewisian semantics to a higher dimension, which is in turn victimized by similar counter-examples, and so on, as we suggest, for any (finite) dimension.

In §2 we explain the consequence argument in more detail, including Pruss’s formal proof of BETA-2; in §3 we present counter-examples to BETA-2 formulated with the actuality operator as well as with rigidified descriptions; in §4 we revise the Lewisian semantics for counterfactuals in a way that avoids the counter-examples, formulate a novel “no-choice-about” principle that transfers across conditionals, and formulate a new consequence argument. Yet, we argue that counter-examples to this novel principle (and many others) can be generated in an interesting way.

## 2 Background

### 2.1 The Consequence Argument

Van Inwagen’s (1983: 56) informal presentation of the consequence argument runs as follows:

---

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither it is up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.

The key of the argument is the assumption that “no-choice about” transfers across conditionals, that is, where  $\varphi$  and  $\psi$  are any statements<sup>1</sup> (possibly about some human action), if  $\varphi$  and no one has or ever had any choice about whether  $\varphi$ ,  $\varphi \rightarrow \psi$  and no one has or ever had any choice about whether  $\varphi \rightarrow \psi$ , then  $\psi$  and no one has or ever had any choice about whether  $\psi$ . This is what the (in)famous rule BETA says in the following standard modal formulation of the consequence argument.<sup>2</sup> Let  $N\varphi$  abbreviate the statement that  $\varphi$  and no one has or ever had any choice about whether  $\varphi$  (cf. van Inwagen 1983: 93-95). Then BETA can be formulated more precisely as follows:

$$\text{BETA } N\varphi, N(\varphi \rightarrow \psi) \vdash N\psi$$

In addition to BETA, the argument assumes another rule of inference involving  $N$ , namely, ALPHA, that tells us that if a statement is necessarily true, then no one has or ever had any choice about it. More formally:

$$\text{ALPHA } \Box\varphi \vdash N\varphi$$

Now, it is a consequence of determinism that any true statement  $\varphi$  (say, that Nina raises her hand) is strictly implied by the laws of nature,  $\Lambda$ , and a statement,  $\Phi$ , describing the complete state of the world in the remote past, that is:

$$\Box((\Lambda \wedge \Phi) \rightarrow \varphi)$$

Then van Inwagen’s modal formulation of the consequence argument is as

---

<sup>1</sup>A statement here is simply a disambiguated declarative sentence.

<sup>2</sup>Although we follow Pruss (2013) in calling the rules governing  $N$  in the consequence argument ALPHA and BETA, van Inwagen (1983: 94) calls them  $\alpha$  and  $\beta$ .

---

follows:

1	$\Box((\Lambda \wedge \Phi) \rightarrow \varphi)$	
2	$N\Lambda$	
3	$N\Phi$	
4	$\Box(\Lambda \rightarrow (\Phi \rightarrow \varphi))$	1
5	$N(\Lambda \rightarrow (\Phi \rightarrow \varphi))$	ALPHA, 4
6	$N(\Phi \rightarrow \varphi)$	BETA, 2, 5
7	$N\varphi$	BETA, 3, 6

Line 4 derives from line 1 by replacing  $(\Lambda \wedge \Phi) \rightarrow \varphi$  with the truth-functionally equivalent  $\Lambda \rightarrow (\Phi \rightarrow \varphi)$  inside the scope of  $\Box$ . So, the argument concludes that if determinism is true, no one has or ever had a choice about  $\varphi$ . McKay and Johnson (1996) provided a decisive counter-example to BETA (which we will not go through here), but Widerker (1987) and Finch and Warfield (1998) suggested another rule in its place which is immune to McKay and Johnson's counter-example besides providing a more elegant and simpler formulation of the consequence argument. This rule is the following:

$$\text{BETA-2 } N\varphi, \Box(\varphi \rightarrow \psi) \vdash N\psi$$

Armed with BETA-2 the incompatibilist can then formulate a new consequence argument:

1	$\Box((\Lambda \wedge \Phi) \rightarrow \varphi)$	
2	$N(\Lambda \wedge \Phi)$	
3	$N\varphi$	BETA-2, 1, 2

This new version of the consequence argument has a nice feature: it does not rely on ALPHA, but only on BETA-2. So whether the new consequence argument is a valid one turns out to be a question about the validity of BETA-2.

In fact, this is a *desideratum* in Huemer's (2000: 529-530) list for a successful consequence argument, that is: (i) the premises of the argument ought to be true, (ii) the claim that  $N\varphi$  holds of all truths  $\varphi$  about human times is incompatible with us having free will, and (iii) BETA-2 must be valid.

---

Now, any claim to validity should be made in the context of a formal background, for even though BETA-2 has the appearance of a reliable and general rule of inference, there is no way to know whether it holds in *every* case without a proof. If we cannot know that, then the new version will not satisfy (iii), and so the new consequence argument will not be successful.

To this end, Pruss (2013) made a significant contribution to the literature by providing a formal interpretation of the  $N$  operator and a proof of the validity of BETA-2 against the background of a widely accepted modal semantics for counterfactuals, namely, Lewis’s (1973) semantics.<sup>3</sup> We now turn to Pruss’s formal definition of  $N$  and his proof of the validity of BETA-2.

## 2.2 Pruss’s proof of BETA-2

Pruss shows that given the standard Lewisian semantics for counterfactuals, Huemer’s condition (iii) is satisfied. He provides the following definition of  $N$  that is intended to capture formally the more intuitive “no-choice-about” condition according to what Huemer calls the “counterfactual interpretation” of  $N$ :<sup>4</sup>

$$N\varphi =_{df.} \varphi \wedge \neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\varphi)]$$

In the definition above,  $\Box\rightarrow$  is the subjunctive conditional or counterfactual,  $x$  is a variable ranging over humans,  $\alpha$  is a variable ranging over all possible action types in the past, present, and future (cf. Pruss (2013: 433)), while  $\text{Can}(x, \alpha)$  and  $\text{Does}(x, \alpha)$  are left unanalysed. This formal definition of  $N$  is indeed appropriate to capture what van Inwagen (1983, especially pp. 66-68) intended in claiming that no one has a choice about  $\varphi$  just in case no one is able to render it false. It is also widely assumed by participants of the debate.

Now recall that in the standard Lewisian semantics for counterfactuals, a counterfactual of the form  $\varphi \Box\rightarrow \psi$  is true (say, at the actual world) if and only if (i) there is no possible world at which  $\varphi$  is true, or (ii)  $\psi$  is true at the closest (or more similar) possible worlds (to the actual world) at which  $\varphi$  is true.<sup>5</sup> Then Pruss observes that according to the Lewisian semantics the following rule of inference, called WEAKENING, is valid:

---

<sup>3</sup>Although, as we shall see, Pruss’s proof is syntactic in nature; but the rules used in the proof can all be shown to be sound according to a (modal) Lewisian semantics for counterfactuals.

<sup>4</sup>A small observation: throughout his paper, Pruss uses the letters  $p, q, r$ , etc., to stand in place of arbitrary statements, whereas we use letters from the end of the Greek alphabet.

<sup>5</sup>What we make, precisely, of the notion of closeness of worlds is orthogonal to the main issue here, so one may simply take Lewis’s (1973: 48) minimal conditions for comparative similarity, for example.

---

WEAKENING  $\varphi \Box \rightarrow \psi, \Box(\psi \rightarrow \chi) \vdash \varphi \Box \rightarrow \chi$

With this rule available, Pruss shows that under the plausible assumption that metaphysical necessity obeys the modal logic **T**, BETA-2 is valid. The proof is offered in Fitch-style natural-deduction rather than by model-theoretic means—this is not an issue, for a model-theoretic proof can be provided within a two-sorted language with two first-order quantifiers, one ranging over individuals, and another over action types.

We reproduce Pruss’s proof below for the reader’s convenience:<sup>6</sup>

---

<sup>6</sup>Pruss uses “taut con” (see the justification for lines 7 and 17) as a (convenient) rule for tautological consequences. We follow him in using this rule in the Fitch-style proof found in §4.1.

---

1	$N\varphi$	
2	$\Box(\varphi \rightarrow \psi)$	
3	$\varphi \wedge \neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\varphi)]$	def, 1
4	$\neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\varphi)]$	$\wedge$ -elim, 3
5	$\forall x\forall\alpha\neg[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\varphi)]$	De Morgan, 4
6	$ba \quad \neg[\text{Can}(b, a) \wedge (\text{Does}(b, a) \Box\rightarrow \neg\varphi)]$	$\forall$ -elim, 5
7	$\text{Can}(b, a) \rightarrow \neg(\text{Does}(b, a) \Box\rightarrow \neg\varphi)$	taut con, 6
8	$\Box[(\varphi \rightarrow \psi) \rightarrow (\neg\psi \rightarrow \neg\varphi)]$	modal contraposition
9	$\Box(\varphi \rightarrow \psi) \rightarrow \Box(\neg\psi \rightarrow \neg\varphi)$	$\Box$ -distr, 8
10	$\Box(\neg\psi \rightarrow \neg\varphi)$	$\rightarrow$ -elim, 2, 9
11	$\quad \quad \quad \text{Can}(b, a)$	
12	$\quad \quad \quad \quad \quad \text{Does}(b, a) \Box\rightarrow \neg\psi$	
13	$\quad \quad \quad \quad \quad \text{Does}(b, a) \Box\rightarrow \neg\varphi$	weakening, 10, 12
14	$\quad \quad \quad \quad \quad \neg(\text{Does}(b, a) \Box\rightarrow \neg\varphi)$	$\rightarrow$ -elim, 7, 11
15	$\quad \quad \quad \quad \quad \neg(\text{Does}(b, a) \Box\rightarrow \neg\psi)$	reductio, 12–14
16	$\text{Can}(b, a) \rightarrow \neg(\text{Does}(b, a) \Box\rightarrow \neg\psi)$	$\rightarrow$ -intro, 11–15
17	$\neg[\text{Can}(b, a) \wedge (\text{Does}(b, a) \Box\rightarrow \neg\psi)]$	taut con, 16
18	$\forall x\forall\alpha\neg[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\psi)]$	$\forall$ -intro, 17
19	$\neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\psi)]$	De Morgan, 18
20	$\varphi \rightarrow \psi$	<b>T</b> , 2
21	$\varphi$	$\wedge$ -elim, 3
22	$\psi$	$\rightarrow$ -elim, 20, 21
23	$\psi \wedge \neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \neg\psi)]$	$\wedge$ -intro, 19, 22
24	$N\psi$	def, 23

Given his proof, Pruss claims that any counter-example to BETA-2 would also have to be a counter-example to WEAKENING and, therefore, to Lewis's semantics for counterfactuals. This seems correct *in a sense*, for WEAKENING is indeed essential for Pruss's proof of the validity of BETA-2. The focus of the debate about the consequence argument, Pruss claims, should then move on from BETA-2 and be about whether  $N(\Lambda \wedge \Phi)$  is really true. For even though

---

many accept that we have no choices about the laws and the remote past, Lewis (1981), for instance, offered arguments against it.

However, there is a sense in which one could still be warranted and well-motivated in rejecting BETA-2 even in light of Pruss’s proof. For Pruss shows that BETA-2 holds relative to an intensional semantics for counterfactuals and the “no-choice-about” operator  $N$ . But many have recently turned to a hyperintensional semantics for counterfactuals, according to which counterfactuals with impossible antecedents fail to be vacuously true as in the orthodox Lewisian semantics.<sup>7</sup> In the case of a hyperintensional semantics, WEAKENING can indeed fail for some impossible antecedent in the first counterfactual displayed in the rule above. But, as noted by Pruss, one could still claim that the regular counterfactual conditional obeys WEAKENING, and that the rule would only fail in the case of an impossible antecedent  $\varphi$ , which is in turn entirely irrelevant for the validity of BETA-2 since  $N\varphi$  forces  $\varphi$  to be true, and therefore  $\psi$  cannot be false, as the argument assumes that  $\Box(\varphi \rightarrow \psi)$  also holds.

We think this is correct. But there is yet another possibility, which does not involve non-trivial counterfactuals, at least directly. It follows from BETA-2 that if a statement is necessarily true, no one has or ever had a choice about it. For let  $\psi$  be necessary, that is, true at every possible world. Then  $\psi$  is strictly implied by any statement. Thus, in particular, if  $\varphi$  is a statement such that  $N\varphi$  is true,  $\varphi$  strictly implies  $\psi$ , that is,  $\Box(\varphi \rightarrow \psi)$  is true. And hence  $N\psi$  is true, too, according to BETA-2. So the rule excludes the possibility of us having choices about necessary truths: if a statement is necessarily true, it just follows from BETA-2 that no one has or ever had a choice about it—which is what is stated by ALPHA after all.

This is a central tenet of the rules used in the consequence argument and, furthermore, of the assumption that having a choice about a true statement is to be able to render it false (where the latter notion is understood counterfactually). But it is not obvious to us that this is correct. In fact, we believe there are counter-examples to the claim that no one has or ever had a choice about a necessary truth. And if this is right, the background semantics for  $N$  should take into account hyperintensional contexts of *some kind*, for there would be necessarily equivalent statements  $\varphi$  and  $\psi$  such that no one has or ever had a choice about  $\varphi$  while the same is not true about  $\psi$ .

As the next section makes apparent, these cases are not directly related to the hyperintensional movement against the orthodox Lewisian semantics.

---

<sup>7</sup>See for instance, Nolan (1997, 2013), Brogaard and Salerno (2007, 2013), and Berto et al. (2018).



---

Rather, they involve more traditional pieces of modal apparatus, to wit, the actuality operator and rigidified descriptions.

### 3 Reasons for rejecting BETA-2

One of the most interesting aspects of van Inwagen’s modal formulation of the consequence argument is that—witnessed by Pruss’s proof—it has brought the modal machinery of possible worlds semantics to the center of one of the most important debates concerning free will. Yet, we believe that an important piece of modal machinery, namely, the actuality operator, might cause troubles for the incompatibilist after all.

First recall the truth-conditions for the actuality operator. Relative to a Kripke model  $\mathfrak{M}$  defined as usual for the basic propositional modal language with the addition of the actuality operator, @, @ $\varphi$  is actually true at a possible world  $w$  if and only if  $\varphi$  is true at the actual world of the model, say,  $w^*$ . More formally:

$$\mathfrak{M}, w \models @\varphi \iff \mathfrak{M}, w^* \models \varphi$$

Now let  $\varphi$  be the statement, say, that Nina raises her hand. Then someone has a choice about whether  $\varphi$  is true, namely, Nina herself. Moreover, if Nina actually exists, she has a choice about whether  $\varphi$  is true *at the actual world*. But from this fact it is reasonable to conclude that she also has a choice about whether the actualized statement @ $\varphi$  is true at the actual world, even though @ $\varphi$  and  $\varphi$  have different modal profiles: the former is necessary if true, whereas the latter is only contingent.

Why do we believe that Nina has a choice regarding the truth of @ $\varphi$  at the actual world? Well, we believe this because the truth of both @ $\varphi$  and  $\varphi$  obtain at the actual world for the very same reasons, that is, the very same facts, grounds, states of affairs, truthmakers or what have you, that make @ $\varphi$  true at the actual world are the ones that make  $\varphi$  true at the actual world. After all, @ $\varphi$  is true at the actual world *in virtue of* Nina’s raising her hand in the actual world, and the same holds of  $\varphi$ . It was Nina’s decision to raise her hand in the actual world, thereby making both @ $\varphi$  and  $\varphi$  true. In a nutshell: the truth of both statements at the actual world is up to her. Of course there are possible worlds wherein @ $\varphi$  is true and  $\varphi$  false, namely, any other world wherein Nina does not raise her hand. If, however, we are concerned about choices made in the actual world, then we could treat the differing modal profiles of @ $\varphi$  and  $\varphi$  as an irrelevant difference.

---

But if one thinks, as we do, that Nina has a choice about the truth of  $@\varphi$  in the actual world it is easy to see why one should also reject BETA-2. For if  $@\varphi$  is true, it is strictly implied by any statement, since actualized statements are necessarily true if true at all. So let  $\psi$  be any statement such that  $N\psi$  holds. Then even though both  $N\psi$  and  $\Box(\psi \rightarrow @\varphi)$  are true, we think that  $N@\varphi$  can be false, and therefore that BETA-2 is invalid.

Note, however, that if  $@$  is added to the language in which Pruss’s proof was formulated, the proof still goes through. As a matter of fact, it is simple to see that according to Pruss’s definition of  $N$ , for any true statement of the form  $@\varphi$ ,  $N@\varphi$  is true, too, since for any possible world  $w$ , human  $x$ , action type  $\alpha$ , if  $\text{Does}(x, \alpha)$  is true at  $w$ , then  $@\varphi$  is also true at  $w$ , since  $@\varphi$  is true at every possible world. Yet, we have argued that some statements of the form  $@\varphi$  are up to us, and therefore that the counterfactual sufficiency interpretation of “no-choice-about” get things wrong in the context of actualized statements, for no one would be able to render these false by acting otherwise according to the formal interpretation of  $N$ . This is why we say that these cases are *intuitive*—as opposed to *formal*—counter-examples to BETA-2.<sup>8</sup>

There are multiple ways one could object to this. One could defend that the explanation or grounds for the truth of a statement  $\varphi$  should be the same at every possible world wherein it is true. And since there can be worlds where  $\varphi$  is false but  $@\varphi$  true, the explanation for the truth of these statements must vary across worlds. However, the claim that the explanation for the truth of a statement should be the same at every world where that statement is true is not plausible. A simple disjunction  $\varphi \vee \psi$  might be true at several possible worlds, and yet its truth can be grounded in different facts across worlds: at a world  $w$  the disjunction is true in virtue of  $\varphi$  being true, and at a world  $w'$  in virtue of  $\psi$  being true, for example.<sup>9</sup>

Another objection involves the fact that we are taking the explanation for the truth of a statement to be bounded, in a sense, by a possible world. For we say that the explanation for the truth of both  $\varphi$  and  $@\varphi$  *at the actual world* is the same, rather than the more general claim that the explanation for the truth of both  $\varphi$  and  $@\varphi$  is the same. If, however, the explanandum finds

---

<sup>8</sup>We should note that Kearns (2011) has independently argued that the actuality operator creates trouble for the incompatibilist in the context of the direct argument. Of course, we agree with him, and the present arguments can be seen as a generalization of his case against the direct argument.

<sup>9</sup>Note, additionally, that our cases are compatible with—although not dependent upon—the *necessitation of grounding*, which is widely held (see Witmer et al. (2005), Rosen (2010), Trogdon (2013), Dasgupta (2014), and many others). This is the thesis that if the fact  $[\varphi]$  grounds a fact  $[\psi]$ , then  $[\psi]$  is necessitated by  $[\varphi]$ .

---

itself within a modal context, we think our account is exactly as it should be. For if we are considering modal contexts where statements are true or false at different possible worlds it makes no sense to say that a statement is simply true, without reference to the world at which it is true. And so, if we are considering what explains the truth of a certain statement at some point in the modal space, we ought also to relativize it to that very point.

Yet another objection would involve rejecting the actuality operator altogether by claiming, for instance, that it does not correspond to the natural language usage of “actually”. We think this objection misses the mark though. Sure, some have argued that @ does not correspond to the way we use “actually” in English.<sup>10</sup> But the main point here has nothing to do with how and whether @ applies to natural language discourse. For there is obviously a widespread technical use of @ that serves many philosophical purposes, including certain developments in epistemic two-dimensionalism (Chalmers (2004, 2006)) and context dependence more generally (Davies and Humberstone (1980), Davies (1983, 2004), Percival (1991)), whether counterpart theory provides the best way to understand modal discourse (Fara and Williamson (2005)), the semantics for indicative conditionals (Weatherston (2001)), and so on.

Still, the main point against BETA-2 can be made even without explicit reference to the actuality operator, for there is yet another piece of modal machinery that can be used in order to argue against BETA-2, namely, rigidified descriptions. For suppose that in the early days of The Beatles it was undecided who would be the lead singer, as John and Paul would more often than not share the burden during rehearsals. At one point John volunteered, deciding it would be best if he was the lead singer; a decision which was promptly accepted by Paul and the others.<sup>11</sup> Then it is true that the lead singer of The Beatles is John Lennon, even though this is contingent. For there are possible worlds where, say, the decision to take on the microphone was made by Paul rather than John. Yet, consider the rigidified description *the (actual) lead singer of The Beatles*. This picks out John Lennon at every possible world, since it picks out John Lennon at the actual world in the first place. And it does this precisely in virtue of the fact that John Lennon made the decision to be the lead singer of The Beatles. In this case, the statement that the (actual) lead singer of The Beatles is John Lennon is necessarily true, even though this was up to John in the sense just explained. And because this statement is

---

<sup>10</sup>See Yalcin (2015). In particular, Yalcin rejects the claim that there are two semantic roles performed by “actually”, namely, a logical and a rhetorical one.

<sup>11</sup>Pretend that in this alternate world Paul was having troubles singing and playing the bass simultaneously.

---

necessarily true, it is strictly implied by any statement, including any statement  $\varphi$  such that  $N\varphi$  is true. Therefore, rigidified descriptions also provide counter-examples to BETA-2.<sup>12</sup>

## 4 A two-dimensional semantics for counterfactuals

Since there are some necessary truths that are up to us, the notion of choice about a true statement cannot involve the ability to render it false, since no one can render a necessary truth false. But in order to salvage the intuition that having a choice about a true statement is to be able to render it false in light of the above cases, one may want to capture formally the intuitive idea according to which, had the actual world been different, actualized statements would be false.

For instance, if it is true at the actual world that Nina raises her hand, then this is actually true, and so it is necessary that Nina actually raises her hand. Yet, had the actual world itself been different, say, a world in which Nina refrains from raising her hand, then it would not be true (at any world) that Nina actually raises her hand after all. Thus, even though when it is true, the statement that Nina actually raises her hand is a necessary truth, the thought continues, there remains a sense in which it seems to be contingent: it does not appear to be inevitable that the actual world turned out in a way that made it true. The Lewisian semantics (or even modal semantics, more generally) for subjunctive conditionals is, however, clearly insufficient to offer a formal basis for this idea.

---

<sup>12</sup>We should note that our cases also provide compelling reasons for rejecting the consequence argument formulated with Pruss’s  $M$  operator, which in turns corresponds to a more formal rendering of Huemer’s (2000)  $N_S$  operator:

$$M\varphi =_{df.} \varphi \wedge \forall x\forall\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \Box\rightarrow \varphi)]$$

This is supposed to capture the idea that  $\varphi$  is true no matter what. Yet, even though the statements that Nina actually raises her hand and that the (actual) lead singer of The Beatles is John Lennon are true no matter what in the sense above, they were up to Nina and John. So, again, even though Pruss’s proof of the validity of GAMMA-2, that is,

$$\text{GAMMA-2 } M\varphi, \Box(\varphi \rightarrow \psi) \vdash M\psi$$

still holds if the formal language is enriched with an actuality operator—or rigidified descriptions—we think the rule is intuitively invalid, for it validates inferences that we would otherwise reject.

---

But what if a subjunctive conditional were to take into account not only the closest worlds where its antecedent is true, but the closest worlds *considered as actual* where its antecedent is true? That is to say, in evaluating a subjunctive conditional *if  $\varphi$  had been the case, then  $\psi$  would have been the case*, whenever we consider a possible world that is relevantly similar to the actual world at which  $\varphi$  is true, we take that very world as if it were actual, and so the actual world also varies in the evaluation of subjunctive conditionals.<sup>13</sup> This has the consequence that even if  $@\varphi$  is true at the actual world, there can be a possible world  $w$  such that, if  $\varphi$  is false at  $w$ , then  $@\varphi$  is false at  $w$  taken as actual. This new semantics would provide a way to falsify true statements of the form  $@\varphi$  at different possible worlds, which will in turn be useful as a suggestion of how one can render false those necessary truths.<sup>14</sup>

For the purpose of illustration, consider a subjunctive conditional *if  $\varphi$  had been the case, then  $\psi$  would have been the case* that is true in the actual world. To mark the difference between the proposed semantics for subjunctive conditionals and the Lewisian semantics, let us use the symbol  $\blacksquare \rightarrow$  for the subjunctive conditional under the proposed semantics, and reserve the symbol  $\square \rightarrow$  for the subjunctive conditional under the Lewisian semantics.<sup>15</sup> Now let  $w_0$  be the actual world, and consider a sequence of possible worlds,

$$w_0, w_1, w_2, w_3, \dots$$

Let the worlds  $w_1, w_2$ , and  $w_3$  be the closest worlds to  $w_0$  where  $\varphi$  is true. Then even though  $\varphi$  might be true in every other world up to  $w_n$ , say, in case the sequence is finite (it does not matter if it is), let these be sufficiently

---

<sup>13</sup>A similar idea is defended by Weatherson (2001) and Wehmeier (2013) for indicative conditionals, where the change of actual world reflected an epistemic feature of indicatives, instead of the metaphysical feature of subjunctives. Yet, these semantics do not assume that there is a relation of similarity for indicatives, in which case an indicative conditional really quantifies over every possible world (in a model) taken as actual.

<sup>14</sup>Another feature of this semantics is that it provides an account of why—regardless of their agential powers—agents cannot falsify some contingent a priori truths, i.e. contingent truths that are true at every possible world  $w$  from the perspective of  $w$  as the actual world (in other words, contingent truths that hold at every diagonal point). We should note that Fusco (2019) explores this topic in her account of some of the puzzles of deontic logic by making use of a two-dimensional semantics involving a diagonal consequence relation, which relates to the present project given the strong connection between free will and moral responsibility.

<sup>15</sup>More generally, we should speak of the semantics for subjunctive conditionals developed by Stalnaker (1968), Lewis (1973), and Kratzer (1979). Even though there are differences amongst these, the general idea is the same. But since Pruss proves the validity of BETA-2 in the Lewisian semantics, this is the one we consider here.

---

dissimilar to the actual world and therefore ignored by the conditional. Since we are considering formulas to be true at a world from the perspective of that world as the actual one, formulas will be evaluated with respect to a pair of worlds, namely, the usual world of evaluation, and the actual world in question. Then we can represent how the conditional  $\blacksquare \rightarrow$  works with a matrix where the worlds distributed horizontally are worlds of evaluation and the worlds distributed vertically are the very same worlds taken as actual. So  $\blacksquare \rightarrow$  considers the closest worlds  $w_i$  to  $w_0$  where  $\varphi$  is true with  $w_i$  taken as actual, which are the points marked with the black dots in the diagonal of the matrix below:

	$w_0$	$w_1$	$w_2$	$w_3$	$\dots$
$w_0$	•				$\dots$
$w_1$		•			$\dots$
$w_2$			•		$\dots$
$w_3$				•	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

We call this kind of conditionals *counteractuals*, to distinguish it from the orthodox counterfactual conditional interpreted with the Lewisian semantics.<sup>16</sup>

As will be made clear below, counteractuals are by and large inspired in the two-dimensional semantics offered in Davies and Humberstone (1980) to distinguish formally between two notions of necessity, namely, superficial and deep necessity, which were in turn informally developed by Evans (1979) to account for Kripke’s (1980) examples of the contingent a priori.<sup>17</sup>

---

<sup>16</sup>It is more proper to use “counterfactuals” for subjunctive conditionals with false antecedents, but we pass over this distinction here since it is not important in what follows.

<sup>17</sup>Our use of the term “counteractual” is similar—though not quite the same, we think—to Yablo’s (2002) original use. Rather than considering a world  $w$  as counterfactual, Yablo talks about considering a world  $w$  as counteractual, i.e. as a “hypothesis about what *this* world is like.” (2002: 449). This is done as a proposal to understand the notion of conceptual necessity and its differences with respect to metaphysical necessity: the former would consider worlds as counteractuals, whereas the latter as counterfactuals. Yet, we are not trying to draw a distinction between conceptual and metaphysical necessity here, for this would bring obvious questions about how we should think of the reference of names in counteractual worlds, and the use of language more broadly, all of which is beyond the scope

---

In the formalism offered by Davies and Humberstone, superficial necessity is the notion formalized by  $\Box$ , whereas deep necessity is formalized by appending the actuality operator to the right of the fixedly operator,  $\mathcal{F}$ , which quantifies universally over possible worlds taken as actual. So  $\mathcal{F}@$  $\varphi$  says that no matter which world turns out to be actual,  $\varphi$  would have been true there. Although we agree that this is a genuine notion of necessity, we do not need to commit to Davies and Humberstone’s claim that  $\mathcal{F}@$  has something to do with a priori knowability, although this would be a standard reading of  $\mathcal{F}@$  as a diagonal operator. In addition, note that we cannot formalize counterfactuals as  $\mathcal{F}@\varphi \rightarrow \psi$ , as this would imply that  $\varphi \rightarrow \psi$  is true at *every* world taken as actual, whereas what we want is just that  $\psi$  be true at the closest worlds to the actual one where  $\varphi$  is true.<sup>18</sup>

To be more precise, we now give a brief outline of the syntax and semantics of a formal language **CAL**+ $\Box$ + $@$  for counterfactual conditionals with distinct modal operators.<sup>19</sup> Then let **PROP** be a denumerably infinite set of sentential letters  $p, q, r, \dots$ , and let  $\neg$  stand for negation,  $\wedge$  for conjunction,  $\blacksquare \rightarrow$  for the counterfactual conditional,  $@$  be the actuality operator,  $\Box$  the counterfactual necessity operator, and  $\blacksquare$  the counterfactual necessity operator—the distinction between  $\Box$  and  $\blacksquare$  will be made clear once the semantics is presented. Intuitively,  $@$  is read as “it is actually the case that...”,  $\Box$  is read as “it is necessary that...”, and  $\blacksquare$  is read as “no matter which world is actual, it is actually

---

of this paper. We should also note that Stefánsson (2018) develops a multidimensional semantics for counterfactuals to avoid what has been called *counterfactual skepticism*, i.e. (roughly) the view that all counterfactuals are false. In his framework, however, the word “counterfactual” stands for something else: a possibly counterfactual world is a truth-maker for a counterfactual statement (cf. Stefánsson (2018: 885)).

<sup>18</sup>One might object that the semantics for  $\blacksquare \rightarrow$  has an epistemic feature, since it quantifies over worlds taken as actual. But we do not think this conditional ought to be taken as epistemic. The notion of necessity formalized by  $\mathcal{F}@$ , for example, even though it has epistemic applications, is not epistemic, as Davies and Humberstone are clear about the fact that they are not offering an epistemic logic. Moreover, we should point out that the fixedly operator itself appeared earlier in Crossley and Humberstone (1977), with motivations that are similar to ours. When Crossley and Humberstone axiomatized the modal logic of actuality they noted that some—although they were not committed with this claim—might think the axiom schema  $@\varphi \rightarrow \Box@\varphi$  counter-intuitive for the reason that it might seem contingent in some sense which world turned out to be actual. Because  $@\varphi \rightarrow \mathcal{F}@\varphi$  is not valid in the modal logic of fixedly actually, the necessity formalized by  $\mathcal{F}@$  might then be taken as the necessity one has in mind in denying that actual facts are all necessarily actual. The application to a priori knowledge, however, was suggested only later in Davies and Humberstone (1980) in connection to Evans’s work.

<sup>19</sup>The language here is propositional and so it does not contain rigidified descriptions. The main philosophical points we will be making here carry over to the case of rigidified descriptions since any non-rigid description can be rigidified by using the actuality operator.

---

the case that...”. The language  $\mathbf{CAL} + \Box + @$  is then recursively generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \blacksquare \rightarrow \varphi) \mid @\varphi \mid \Box\varphi \mid \blacksquare\varphi$$

The other Boolean connectives such as disjunction  $\vee$  and the material conditional  $\rightarrow$  are defined as usual.<sup>20</sup>

A *model*  $\mathfrak{M}$  for  $\mathbf{CAL} + \Box + @$  is a tuple,  $\mathfrak{M} = (W, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , where  $W$  is a non-empty set of possible worlds,  $\{R_\varphi \mid \varphi \in \mathbf{F}\}$  is a set of binary (“accessibility”) relations on  $W$ , one for each formula  $\varphi \in \mathbf{F}$ , where  $\mathbf{F}$  is the set of formulas of  $\mathbf{CAL} + \Box + @$ ,<sup>21</sup> and  $V$  is a function assigning the value 1 or 0 to each  $p \in \mathbf{PROP}$  at a possible world  $w \in W$ , written  $V_w(p) = 1$  (or 0).

This way of defining models for conditional languages is based on Priest (2008: ch. 5), although one can find the main ideas much earlier in Chellas (1975). Lewis (1973), by contrast, makes use of so-called systems of spheres, where each  $w \in W$  has its own system of similarity spheres, that is, a set  $\$w$  of sets  $S, S', \dots$  of worlds each of which contains worlds that are more similar to  $w$  than any other world outside of that sphere. We choose Priest’s formulation merely because of its simplicity, but the semantics given here could be defined by means of Lewis’s systems of spheres, too.<sup>22</sup>

The relations  $R_\varphi$  can then be thought of as representing similarity, that is, where  $wR_\varphi w'$  is taken intuitively to mean that  $w'$  is a world most similar to  $w$  such that  $\varphi$  is true, or what Priest (2008: 84) calls a “ceteris paribus” condition, where  $wR_\varphi w'$  in turn means that  $w'$  is ceteris paribus the same as  $w$ , but  $\varphi$  is true at  $w'$ .

Let us write  $\mathfrak{M}, w, v \models \varphi$  for “ $\varphi$  is true at world  $w$  with respect to  $v$  taken as the actual world in  $\mathfrak{M}$ ”. Now the notion of truth at a world with respect to a world taken as actual is extended to all formulas recursively relative to a

---

<sup>20</sup>We drop the outer parentheses of formulas when convenient.

<sup>21</sup>That is,  $\mathbf{F} = \mathbf{CAL} + \Box + @$ . The reason why we use  $\mathbf{F}$  is because there are models which we define below with a set of accessibility relations, one such relation for each formula, again, but over a different language. So  $\mathbf{F}$  makes the notation of these models simpler.

<sup>22</sup>A third possibility is to use selection functions, as in Chellas (1975) and Priest (2008: ch. 5).



---

model  $\mathfrak{M}$  for **CAL** +  $\square$  +  $@$  as follows:

$$\begin{array}{ll}
\mathfrak{M}, w, v \models p & \iff V_w(p) = 1 \\
\mathfrak{M}, w, v \models \neg\varphi & \iff \mathfrak{M}, w, v \not\models \varphi \\
\mathfrak{M}, w, v \models (\varphi \wedge \psi) & \iff \mathfrak{M}, w, v \models \varphi \text{ and } \mathfrak{M}, w, v \models \psi \\
\mathfrak{M}, w, v \models (\varphi \blacksquare \rightarrow \psi) & \iff \text{for all } w' \in W, \text{ if } wR_\varphi w', \text{ then } \mathfrak{M}, w', w' \models \psi \\
\mathfrak{M}, w, v \models @\varphi & \iff \mathfrak{M}, v, v \models \varphi \\
\mathfrak{M}, w, v \models \square\varphi & \iff \text{for all } w' \in W, \mathfrak{M}, w', v \models \varphi \\
\mathfrak{M}, w, v \models \blacksquare\varphi & \iff \text{for all } w' \in W, \mathfrak{M}, w', w' \models \varphi
\end{array}$$

The truth conditions for the Boolean compounds and for  $@$ - and  $\square$ -formulas are just as usual in two-dimensional modal logics.<sup>23</sup> We have not added a separate accessibility relation for  $\square$ , but this could be done as usual, with different constraints such as reflexivity, seriality, transitivity, etc, generating in turn different modal logics. As it is, the **S5** axioms hold for  $\square$ . The  $\blacksquare$  operator for counteractual necessity corresponds to Davies and Humberstone's  $\mathcal{F}@$ . The new clause is that of the counteractual conditional  $\blacksquare \rightarrow$ . To see how this differs from the truth conditions for counterfactuals, note that in the usual, one-dimensional framework, the truth conditions for counterfactual conditionals would be defined as

$$\mathfrak{M}, w \models (\varphi \square \rightarrow \psi) \iff \text{for all } w' \in W, \text{ if } wR_\varphi w', \text{ then } \mathfrak{M}, w' \models \psi$$

And in the two-dimensional framework, they would be defined as

$$\mathfrak{M}, w, v \models (\varphi \square \rightarrow \psi) \iff \text{for all } w' \in W, \text{ if } wR_\varphi w', \text{ then } \mathfrak{M}, w', v \models \psi$$

That is, while the actual world remains fixed in the definition above, the same is not true for counterfactuals where the actual world varies alongside the counterfactual world.

It is widely known that there are many constraints one can add to the relations  $R_\varphi$  to the end of representing similarity between worlds.<sup>24</sup> One plausible constraint is that for any world  $w'$  that is  $R_\varphi$ -accessible from a world  $w$ ,  $\varphi$  is true at  $w'$  with respect to  $w'$  taken as actual. That is, all the most similar worlds to  $w$ —where  $\varphi$  is true—will be worlds according to which  $\varphi$  is true, too. We can write this first constraint on the relations  $R_\varphi$  as follows: if  $wR_\varphi w'$ , then  $\mathfrak{M}, w', w' \models \varphi$ . Call this condition “antecedent preservation”, or (AP).

Additionally, one may want to add the constraint corresponding to what Lewis (1973) calls “weak centering”, or (WC): if  $\varphi$  holds at a possible world

---

<sup>23</sup>For instance, see Davies and Humberstone (1980: 4-5).

<sup>24</sup>See Lewis (1973: 14) for minimal conditions on systems of spheres.

---

$w$ , then  $w$  is one of the worlds that are most similar to—or  $R_\varphi$ -accessible from— $w$ .<sup>25</sup> More formally, the condition is that if  $\mathfrak{M}, w, v \models \varphi$ , then  $wR_\varphi v$ . There are many other constraints one could impose on the relations  $R_\varphi$ , but this is somewhat orthogonal to our discussion, and so we just assume the minimal conditions (AP) and (WC), which are very plausible for the “similarity” interpretation of the relations  $R_\varphi$ .

We should now define an appropriate notion of logical consequence for **CAL** +  $\square$  +  $\@$ . As per usual with a two-dimensional semantics, there are multiple options on the table. We prefer the notion of *diagonal logical consequence*, which can be defined as follows:

(D-consequence) A formula  $\varphi$  is a *diagonal logical consequence* of a set of formulas  $\Sigma$  if and only if for every model  $\mathfrak{M}$  and world  $w$  in  $\mathfrak{M}$ , if  $\mathfrak{M}, w, w \models \psi$  for all  $\psi \in \Sigma$ , then  $\mathfrak{M}, w, w \models \varphi$ .

Then we say that a formula is *diagonally valid* if it is a diagonal consequence of the empty set. This is the notion of validity (and consequence) one finds in Kaplan (1989), for instance, whereby the formal rendering of statement such as “I am here now” are (diagonally) valid—this statement is true in every proper context of utterance, even though it is also contingent. On the other hand, one could define what Davies and Humberstone (1980: 1) call *general logical consequence* as truth preservation at *all* points in every model. Clearly, diagonal consequence is a special case, and thus entailed by, general consequence. The main reason to adopt diagonal consequence here is in order to validate the **T** axiom schema for  $\blacksquare$ , that is,  $\blacksquare\varphi \rightarrow \varphi$ , which does not hold if validity is defined generally, and it might be reasonably seen as a requirement of any genuine notion of necessity: if it is necessary, in some objective sense, that  $\varphi$ , then  $\varphi$  is true.<sup>26</sup>

## 4.1 The “Counterfactual Sufficiency” Interpretation

Now we can redefine the formal rendering of “no-choice-about” by means of a counterfactual conditional. Inspired by the counterfactual sufficiency interpretation, we call this proposal the “counterfactual sufficiency” interpretation. We use the boldface symbol **N** for this in contrast to the italicized  $N$  that received the counterfactual sufficiency interpretation:

---

<sup>25</sup>This and the previous constraint are added to the (non-vacuumist) semantics for subjunctive conditionals in Berto et al. (2018: 697).

<sup>26</sup>That the **T** axiom schema is not generally valid in the logic of fixedly actually is pointed out in Davies and Humberstone (1980: 6). The culprit here is the actuality operator, for without it **T** is generally valid in the semantics defined here.

---


$$\mathbf{N}\varphi =_{df.} \varphi \wedge \neg\exists x\exists\alpha[\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare\rightarrow \neg\varphi)]$$

This is compatible with Nina’s ability to act otherwise with respect to actually raising her hand. Now consider the statement that Nina actually raises her hand. There is an agent  $x$  (Nina) and action  $\alpha$  (say, refraining from raising her hand) such that  $\text{Can}(x, \alpha)$  and at the closest world  $w$  to the actual world, if  $w$  were the actual world and  $\text{Does}(x, \alpha)$ , then it would be false that Nina actually raises her hand at  $w$ . Therefore,  $\mathbf{N}$  is an appropriate formal rendering of “no-choice-about” in light of our examples. It is still compatible with us having choices about some necessary truths, but it is also consistent with the idea that free will requires the possibility of making a true proposition false.

Can one run a consequence argument with  $\mathbf{N}$ ? Consider the following version of BETA, which we call BETA-3:

$$\text{BETA-3 } \mathbf{N}\varphi, \blacksquare(\varphi \rightarrow \psi) \vdash \mathbf{N}\psi.$$

Then a consequence argument can be simply formulated as follows:

$$\begin{array}{l|l} 1 & \blacksquare((\Lambda \wedge \Phi) \rightarrow \varphi) \\ 2 & \mathbf{N}(\Lambda \wedge \Phi) \\ \hline 3 & \mathbf{N}\varphi \qquad \text{BETA-3, 1, 2} \end{array}$$

To satisfy Huemer’s desiderata (i)–(iii), it remains for us to argue that the premises of this consequence argument are true, and that BETA-3 is valid with respect to this semantics for counterfactuals.

Concerning the validity of BETA-3, it is simple to generalize Pruss’s corresponding proof of BETA-2 to the present case. In Pruss’s proof, the only steps that crucially involve  $\square$  and  $\square\rightarrow$  concern an application of the  $\mathbf{T}$  axiom schema for  $\square$  and the weakening rule for  $\square\rightarrow$  (that is, the rule of inference saying that if both  $\varphi \square\rightarrow \psi$  and  $\square(\psi \rightarrow \chi)$  are true, then  $\varphi \square\rightarrow \chi$  is true). But, as we have observed already, by assuming diagonal validity, the  $\mathbf{T}$  axiom schema for  $\blacksquare$  is valid, and the argument for the validity of weakening for counterfactuals (that is, if both  $\varphi \blacksquare\rightarrow \psi$  and  $\blacksquare(\psi \rightarrow \chi)$  are true, then  $\varphi \blacksquare\rightarrow \chi$  is true) is similar to the argument for the counterfactual case. In fact, weakening is generally (and hence, diagonally) valid:

Suppose that  $\mathfrak{M}, w, v \models \varphi \blacksquare\rightarrow \psi$  and  $\mathfrak{M}, w, v \models \blacksquare(\psi \rightarrow \chi)$ . Then for all  $w' \in W$ , if  $wR_\varphi w'$ ,  $\mathfrak{M}, w', w' \models \psi$ . Suppose that  $z$  is such that  $wR_\varphi z$ . By the truth conditions for  $\blacksquare$ -formulas,  $\mathfrak{M}, z, z \models \psi \rightarrow \chi$ ,

and so  $\mathfrak{M}, z, z \models \chi$  by modus ponens. So, for all  $w' \in W$ , if  $wR_\varphi w'$ ,  $\mathfrak{M}, w', w' \models \chi$ . Therefore, by the truth conditions for  $\blacksquare \rightarrow$ -formulas,  $\mathfrak{M}, w, v \models \varphi \blacksquare \rightarrow \chi$ .

Therefore, Pruss's proof can be simply rewritten by replacing  $\Box$  and  $\Box \rightarrow$  with  $\blacksquare$  and  $\blacksquare \rightarrow$ , respectively, thereby establishing the validity of BETA-3 with respect to the class of models for **CAL** +  $\Box$  + @.

1	$\mathbf{N}\varphi$	
2	$\blacksquare(\varphi \rightarrow \psi)$	
3	$\varphi \wedge \neg \exists x \exists \alpha [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \varphi)]$	def, 1
4	$\neg \exists x \exists \alpha [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \varphi)]$	$\wedge$ -elim, 3
5	$\forall x \forall \alpha \neg [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \varphi)]$	De Morgan, 4
6	$ba \mid \neg [\text{Can}(b, a) \wedge (\text{Does}(b, a) \blacksquare \rightarrow \neg \varphi)]$	$\forall$ -elim, 5
7	$\text{Can}(b, a) \rightarrow \neg (\text{Does}(b, a) \blacksquare \rightarrow \neg \varphi)$	taut con, 6
8	$\blacksquare [(\varphi \rightarrow \psi) \rightarrow (\neg \psi \rightarrow \neg \varphi)]$	modal contraposition
9	$\blacksquare (\varphi \rightarrow \psi) \rightarrow \blacksquare (\neg \psi \rightarrow \neg \varphi)$	$\blacksquare$ -distr, 8
10	$\blacksquare (\neg \psi \rightarrow \neg \varphi)$	$\rightarrow$ -elim, 2, 9
11	$\text{Can}(b, a)$	
12	$\text{Does}(b, a) \blacksquare \rightarrow \neg \psi$	
13	$\text{Does}(b, a) \blacksquare \rightarrow \neg \varphi$	weakening, 10, 12
14	$\neg (\text{Does}(b, a) \blacksquare \rightarrow \neg \varphi)$	$\rightarrow$ -elim, 7, 11
15	$\neg (\text{Does}(b, a) \blacksquare \rightarrow \neg \psi)$	reductio, 12–14
16	$\text{Can}(b, a) \rightarrow \neg (\text{Does}(b, a) \blacksquare \rightarrow \neg \psi)$	$\rightarrow$ -intro, 11–15
17	$\neg [\text{Can}(b, a) \wedge (\text{Does}(b, a) \blacksquare \rightarrow \neg \psi)]$	taut con, 16
18	$\forall x \forall \alpha \neg [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \psi)]$	$\forall$ -intro, 17
19	$\neg \exists x \exists \alpha [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \psi)]$	De Morgan, 18
20	$\varphi \rightarrow \psi$	<b>T</b> , 2
21	$\varphi$	$\wedge$ -elim, 3
22	$\psi$	$\rightarrow$ -elim, 20, 21
23	$\varphi \wedge \neg \exists x \exists \alpha [\text{Can}(x, \alpha) \wedge (\text{Does}(x, \alpha) \blacksquare \rightarrow \neg \psi)]$	$\wedge$ -intro, 19, 22
24	$\mathbf{N}\psi$	def, 23

---

Concerning the premises of the consequence argument itself, we believe that the same reasons that compel one to accept the premises of the original consequence argument apply here. Yet, one might think that while premise 1 formulated with  $\Box$  captures an intuitive version of determinism based on a notion of metaphysical or broadly logical necessity that has intuitive appeal and is widespread in metaphysics, the notion of necessity captured by  $\blacksquare$  is not so intuitive. This type of necessity is the usual diagonal necessity of two-dimensional semantics, which is claimed by Davies and Humberstone (1980) to be aligned with a priori truths.<sup>27</sup>

But what matters here is not whether this formulation of determinism is a priori or not. Rather, it is important that the notion of necessity involved in the argument is appropriate. To this end we observe that  $\Box$  and  $\blacksquare$  in fact coincide for *every* formula *except* the ones containing some occurrence of the actuality operator. More formally, for every formula  $\varphi$  of  $\mathbf{CAL} + \Box + @$  with no occurrences of the actuality operator  $@$ ,  $\Box\varphi \leftrightarrow \blacksquare\varphi$  is diagonally (and generally) valid.<sup>28</sup> So there is at least a formal sense in which these two operators are almost equivalent, differing only with respect to formulas containing occurrences of the actuality operator, which is also in a sense what our examples are capturing intuitively, namely, that we have a choice about certain  $\Box$ -truths but not about any  $\blacksquare$ -truth. In this sense,  $\blacksquare$  captures everything that  $\Box$  does, but it extends it appropriately. The same can be said mutatis mutandis about  $\blacksquare \rightarrow$  and  $\Box \rightarrow$ . If we were to add  $\Box \rightarrow$  to  $\mathbf{CAL} + \Box + @$ ,  $\blacksquare \rightarrow$  and  $\Box \rightarrow$  would only be distinguishable by  $@$ -formulas: for all formulas  $\varphi$  and  $\psi$  of  $\mathbf{CAL} + \Box + @ + \Box \rightarrow$  with no occurrences of the actuality operator  $@$ ,  $(\varphi \Box \rightarrow \psi) \leftrightarrow (\varphi \blacksquare \rightarrow \psi)$  is diagonally (and generally) valid.

In fact, we now prove that there is a sound translation  $t_{\mathbf{CAL}}$  from the *pure counteractual language*  $\mathbf{CAL}$ —i.e. the language of propositional logic plus  $\blacksquare$  and  $\blacksquare \rightarrow$ —onto the *pure counterfactual language*  $\mathbf{CFL}$ —i.e. the language of propositional logic plus  $\Box$  and  $\Box \rightarrow$ —and a translation  $t_{\mathbf{CFL}}$  from  $\mathbf{CFL}$  onto  $\mathbf{CAL}$  such that for all  $\mathbf{CFL}$ -formulas  $\varphi$ , models  $\mathfrak{M}$  for  $\mathbf{CFL}$ , and  $w, v$  in  $\mathfrak{M}$ ,

$$\mathfrak{M}, w, v \models \varphi \iff \mathfrak{M}, w, v \models t_{\mathbf{CAL}}(t_{\mathbf{CFL}}(\varphi))$$

and for all  $\mathbf{CAL}$ -formulas  $\varphi$ , models  $\mathfrak{M}$  for  $\mathbf{CAL}$  and  $w, v$  in  $\mathfrak{M}$ ,

$$\mathfrak{M}, w, v \models \varphi \iff \mathfrak{M}, w, v \models t_{\mathbf{CFL}}(t_{\mathbf{CAL}}(\varphi))$$

First we define a mapping from  $\mathbf{CAL}$ -formulas onto  $\mathbf{CFL}$ -formulas:

---

<sup>27</sup>The idea that the a priori involves some sort of diagonal necessity is defended by many others. See Chalmers (2006) for an overview of a number of distinct formulations of this and other core notions of two-dimensional semantics.

<sup>28</sup>This corresponds to axiom schema ( $\mathcal{F}6$ ) from Davies and Humberstone (1980: 4).

---

**Definition 4.1** (CAL-translation) The *CAL-translation*  $t_{\text{CAL}}$  from **CAL**-formulas onto **CFL**-formulas is defined as follows:

$$\begin{aligned}
t_{\text{CAL}}(p) &= p \\
t_{\text{CAL}}(\neg\varphi) &= \neg t_{\text{CAL}}(\varphi) \\
t_{\text{CAL}}(\varphi \wedge \psi) &= t_{\text{CAL}}(\varphi) \wedge t_{\text{CAL}}(\psi) \\
t_{\text{CAL}}(\varphi \blacksquare \rightarrow \psi) &= t_{\text{CAL}}(\varphi) \square \rightarrow t_{\text{CAL}}(\psi) \\
t_{\text{CAL}}(\blacksquare\varphi) &= \square t_{\text{CAL}}(\varphi)
\end{aligned}$$

Now the following lemma can be established:

**Lemma 4.1** For every **CAL**-formula  $\varphi$ , model  $\mathfrak{M} = (W, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w, v \in W$ ,

$$\mathfrak{M}, w, v \models \varphi \iff \mathfrak{M}, w \models t_{\text{CAL}}(\varphi).$$

*Proof.* The proof is by induction on the complexity of  $\varphi$ . For the base case, let  $\varphi = p$ , then  $\mathfrak{M}, w, v \models p$  iff  $V_w(p) = 1$  iff  $\mathfrak{M}, w \models p$  iff  $\mathfrak{M}, w \models t_{\text{CAL}}(p)$ . Assume the inductive hypothesis, from which the Boolean cases follow straightforwardly. Let  $\varphi = (\psi \blacksquare \rightarrow \chi)$ . Then  $\mathfrak{M}, w, v \models \psi \blacksquare \rightarrow \chi$  iff for all  $w'$ , if  $wR_\psi w'$ , then  $\mathfrak{M}, w', w' \models \chi$  iff (by I. H.) for all  $w'$ , if  $wR_{t_{\text{CAL}}(\psi)} w'$ , then  $\mathfrak{M}, w' \models t_{\text{CAL}}(\chi)$  iff  $\mathfrak{M}, w \models t_{\text{CAL}}(\psi) \square \rightarrow t_{\text{CAL}}(\chi)$ . Let  $\varphi = \blacksquare\psi$ . Then  $\mathfrak{M}, w, v \models \blacksquare\psi$  iff for all  $w'$ ,  $\mathfrak{M}, w', w' \models \psi$  iff (by I. H.) for all  $w'$ ,  $\mathfrak{M}, w' \models t_{\text{CAL}}(\psi)$  iff  $\mathfrak{M}, w \models \square t_{\text{CAL}}(\psi)$ .  $\square$

For the other direction, consider the following mapping from **CFL**-formulas onto **CAL**-formulas, from which a similar lemma can be established:

**Definition 4.2** (CFL-translation) The *CFL-translation*  $t_{\text{CFL}}$  from **CFL**-formulas onto **CAL**-formulas is defined as follows:

$$\begin{aligned}
t_{\text{CFL}}(p) &= p \\
t_{\text{CFL}}(\neg\varphi) &= \neg t_{\text{CFL}}(\varphi) \\
t_{\text{CFL}}(\varphi \wedge \psi) &= t_{\text{CFL}}(\varphi) \wedge t_{\text{CFL}}(\psi) \\
t_{\text{CFL}}(\varphi \square \rightarrow \psi) &= t_{\text{CFL}}(\varphi) \blacksquare \rightarrow t_{\text{CFL}}(\psi) \\
t_{\text{CFL}}(\square\varphi) &= \blacksquare t_{\text{CFL}}(\varphi)
\end{aligned}$$

**Lemma 4.2** For every **CFL**-formula  $\varphi$ , model  $\mathfrak{M} = (W, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w, v \in W$ ,

$$\mathfrak{M}, w \models \varphi \iff \mathfrak{M}, w, v \models t_{\text{CFL}}(\varphi).$$

*Proof.* The proof is by induction on the complexity of  $\varphi$ . For the base case, let  $\varphi = p$ , then  $\mathfrak{M}, w \models p$  iff  $V_w(p) = 1$  iff  $\mathfrak{M}, w, v \models p$  iff  $\mathfrak{M}, w, v \models t_{\text{CFL}}(p)$ . The rest of the proof is analogous to the proof of Lemma 4.1.  $\square$

---

It is simple to observe that the following corollaries now follow from the two lemmas above:

**Corollary 4.1** *For every CAL-formula  $\varphi$ , model  $\mathfrak{M} = (W, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w, v \in W$ ,*

$$\mathfrak{M}, w, v \models \varphi \iff \mathfrak{M}, w, v \models \mathbf{t}_{\text{CFL}}(\mathbf{t}_{\text{CAL}}(\varphi)).$$

**Corollary 4.2** *For every CF-formula  $\varphi$ , model  $\mathfrak{M} = (W, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w \in W$ ,*

$$\mathfrak{M}, w \models \varphi \iff \mathfrak{M}, w \models \mathbf{t}_{\text{CAL}}(\mathbf{t}_{\text{CFL}}(\varphi)).$$

Thus there is a strong sense in which both of these languages are expressively equivalent, but are distinguishable if @ is added to them. That is to say, apart from the formulas containing an actuality operator, one might consider the counterfactual interpretation as just another way of rendering “no-choice-about” formally that says the same thing by different means as the counterfactual interpretation. This offers reason to believe that  $\blacksquare$  and  $\blacksquare \rightarrow$  are not at all less intuitive when compared to the orthodox notions taken to be formalized by  $\Box$  and  $\Box \rightarrow$ . Thus we do not see that there is any disadvantage in formulating the concept of determinism with  $\blacksquare$ . If one feels compelled to accept the premises of the consequence argument when formulated with  $\Box$ , we see no reason for one to reject them when formulated with  $\blacksquare$ .

## 5 Actuality and two-dimensionalism

In this section we argue that despite the apparent success of the counterfactual sufficiency interpretation in light of our examples of necessary truths that are up to us, there are reasons to believe that the move to a two-dimensional semantics might be of no help after all. In a nutshell, new counter-examples can be generated in the two-dimensional case that are similar in spirit to the original counter-examples to BETA-2.

When the actuality operator is added to the basic modal language consisting only of  $\Box$  and  $\Diamond$  besides the usual stock of Boolean connectives we also add a distinguished world to the models that is, intuitively, the actual world of the model. That is, Kripke models in this case are triples,<sup>29</sup>  $\mathfrak{M} = (W, w^*, V)$ , with the truth conditions for @-formulas stated as before. This captures the sense in which a formula is actually true (in a model) if and only if it is true at the actual world (of the model). And this is, in fact, the standard semantic

---

<sup>29</sup>Or quadruples, if one adds an accessibility relation.

---

clause for the actuality operator in this setting.<sup>30</sup> This also reflects the fact that the (world of the) context provided by the model is fixed in the sense that there is no variation of actual world.<sup>31</sup>

Yet, it has become standard practice to present the truth conditions for @-formulas in the two-dimensional setting as we have done it above, that is,

$$\mathfrak{M}, w, v \models @\varphi \iff \mathfrak{M}, v, v \models \varphi$$

Such truth conditions, in turn, reflect the fact that in two-dimensional semantics we can vary the (world of the) context in which a sentence is evaluated. In this case, the actual world is not fixed—recall that models for **CAL** +  $\Box$  + @ do not contain a fixed actual world. So for a sentence to be actually true is just for it to be true relative to the actual world under consideration, which can be any world in the model.<sup>32</sup>

However, one might just as well bring back a distinguished world that is fixed in the models even in the presence of  $\blacksquare$  and  $\blacksquare \rightarrow$ . That is, by letting  $w^*$  be, say, the *really* actual world of the model, the semantic clause for the actuality operator, which in this case we write as **A**, would be given as

$$\mathfrak{M}, w, v \models A\varphi \iff \mathfrak{M}, w^*, w^* \models \varphi$$

There is a question whether @ or **A** (or both) should be the preferred operator for the philosophical notion of actuality captured in the two-dimensional framework. The former is what we want whenever we wish to say that a different world could have been actual or, for instance, when we want to formalize epistemic principles such as *it is a priori that @ $\varphi$  if and only if  $\varphi$* . For this involves conceiving all the ways the actual world could turn out and then verifying whether  $\varphi$  would be true. If the a priori operator receives the semantic clause given for  $\blacksquare$ , then  $\blacksquare(@\varphi \leftrightarrow \varphi)$  is easily seen to be valid, both generally and diagonally, with respect to the models for **CAL** +  $\Box$  + @.  $\blacksquare(A\varphi \leftrightarrow \varphi)$ , on the other hand, is not valid in any sense, as, say,  $p$  might hold at  $w^*$  with respect to  $w^*$  taken as actual, while  $\neg p$  holds at a world  $v$  with respect to  $v$  taken as actual, in which case  $A p \leftrightarrow p$  holds at  $w^*$  with respect to  $w^*$  taken as actual, but  $\blacksquare(A p \leftrightarrow p)$  does not.

---

<sup>30</sup>See, for instance, Hazen (1976), Crossley and Humberstone (1977), Davies and Humberstone (1980), Wehmeier (2004, 2013), to name a few.

<sup>31</sup>We have in mind here Kaplan's (1989) distinction between context of utterance and context of evaluation. The former includes things like the speaker, time, place, and world in which a certain sentence is uttered. The latter provides the circumstances against which that sentence receives a truth-value.

<sup>32</sup>Examples of two-dimensional semantics with these truth conditions for @-formulas include Davies and Humberstone (1980), Restall (2012), Chalmers and Rabern (2014), Fritz (2014), and many others.



---

This situation is perfectly analogous to what happens with  $\Box(@\varphi \leftrightarrow \varphi)$  in the one-dimensional case. For example, while  $@\varphi \leftrightarrow \varphi$  holds at the actual world of any model for the basic modal language with  $@$ , it fails to hold at every world in every model. This is an example of what Zalta (1988) called a contingent logical truth: if logical truth is defined as truth at the actual world of every model, or truth simpliciter,<sup>33</sup> then  $@\varphi \leftrightarrow \varphi$  is a logical truth that is also contingent.<sup>34</sup>

Without taking a stance on the debate about which definition is appropriate for the notion of logical truth in these languages, there are reasons to believe that the notion of actuality captured by the actuality operator in the one-dimensional case is in turn captured by  $\mathbf{A}$ , and not  $@$ , when transported into the two-dimensional framework. For the function performed by  $@$  in the one-dimensional semantics is clearly distinct from the function it performs in the two-dimensional semantics, but it is the same function as the one performed by  $\mathbf{A}$ , namely, to point invariably to the distinguished world of the model. Furthermore, one might just wish to be able to represent different ways the actual world could turn out while keeping a world to represent the really actual world in the models. And, as a matter of fact, we can prove that the translation between CAL and CFL can be extended appropriately to account for  $\mathbf{A}$  and  $@$ :  $\mathbf{CAL} + \mathbf{A}$  and  $\mathbf{CFL} + @$  are also notational variants after all.

Consider the previous mappings but extended with actuality operators as follows:

**Definition 5.1** (CAL\*-translation) The *CAL\*-translation*  $\mathbf{t}_{\mathbf{CAL}^*}$  from  $\mathbf{CAL} + \mathbf{A}$ -formulas onto  $\mathbf{CFL} + @$ -formulas is defined just as the CAL-translation with the addition of the following clause:

---

<sup>33</sup>Corresponding to a generalization of what Crossley and Humberstone (1977) call *real-world validity*.

<sup>34</sup>By transporting this notion of logical truth to the two-dimensional framework, that is, truth at  $w^*$  with respect to  $w^*$  taken as actual in every model, then  $(\mathbf{A}\varphi \leftrightarrow \varphi)$  turns out to be a logical truth that is also counteractually contingent—or deeply contingent, in Davies and Humberstone’s sense. Furthermore, if one takes  $\blacksquare$  as an operator for a priori knowledge, this means that there are logical truths that are not a priori knowable in the sense of being true along the diagonal. That is, besides being able to distinguish the concept of necessity from what is a priori knowable by means of Kripke’s (1980) examples, and necessity from logical truth by means of Zalta’s (1988) examples, the case at hand shows that it is possible to distinguish logical truth from what is a priori knowable, too. In Lampert (2018), this is taken as a possible motivation to call into question the usual interpretation of  $\blacksquare$  as an a priori operator. But, of course, one might simply accept that some logical truths are either a posteriori or unknowable, or even take this as a reductio against this notion of logical truth.

$$t_{\text{CAL}^*}(\mathbf{A}\varphi) = @t_{\text{CAL}^*}(\varphi)$$

Then the following lemma can be easily established:

**Lemma 5.1** *For every CAL + A-formula  $\varphi$ , model  $\mathfrak{M} = (W, w^*, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w, v \in W$ ,*

$$\mathfrak{M}, w, v \models \varphi \iff \mathfrak{M}, w \models t_{\text{CAL}^*}(\varphi).$$

*Proof.* The proof is the same as in Lemma 4.1, except for the new case. Let  $\varphi = \mathbf{A}\psi$ . Then  $\mathfrak{M}, w, v \models \mathbf{A}\psi$  iff  $\mathfrak{M}, w^*, w^* \models \psi$  iff (by I. H.)  $\mathfrak{M}, w^* \models t_{\text{CAL}^*}(\psi)$  iff  $\mathfrak{M}, w \models @t_{\text{CAL}^*}(\psi)$ .  $\square$

Now consider the following extension of the CFL-translation:

**Definition 5.2** (CFL\*-translation) The CFL\*-translation  $t_{\text{CFL}^*}$  from **CFL** + @-formulas onto **CAL** + A-formulas is defined just as the CFL-translation with the addition of the following clause:

$$t_{\text{CFL}^*}(@\varphi) = \text{At}_{\text{CFL}^*}(\varphi)$$

Then, as expected, the following lemma can be established:

**Lemma 5.2** *For every CFL + @-formula  $\varphi$ , model  $\mathfrak{M} = (W, w^*, \{R_\varphi \mid \varphi \in \mathbf{F}\}, V)$ , and  $w, v \in W$ ,*

$$\mathfrak{M}, w \models \varphi \iff \mathfrak{M}, w, v \models t_{\text{CFL}^*}(\varphi).$$

*Proof.* The proof is the same as in Lemma 4.2, except for the new case. Let  $\varphi = @\psi$ , then  $\mathfrak{M}, w \models @\psi$  iff  $\mathfrak{M}, w^* \models \psi$  iff (by I. H.)  $\mathfrak{M}, w^*, w^* \models t_{\text{CFL}^*}(\psi)$  iff  $\mathfrak{M}, w, v \models \text{At}_{\text{CFL}^*}(\psi)$ .  $\square$

Corollaries analogous to 4.1 and 4.2 also follow for the extended languages.

Now, there is strong reason to believe that the explanation for the truth of  $\varphi$  at the really actual world  $w^*$  with respect to  $w^*$  taken as actual is the same as that for the truth of  $\mathbf{A}\varphi$  in the same circumstances. For this mimics what happens in the one-dimensional case with the original actuality operator, @. Therefore the same reasons to believe that one has a choice about certain true instances of @ $\varphi$  in the one-dimensional framework apply for some true instances of  $\mathbf{A}\varphi$  in the two-dimensional framework, which in turn implies that we have choices about counteractual necessities, as  $\mathbf{A}\varphi \rightarrow \blacksquare\mathbf{A}\varphi$  turns out to be

---

valid according to every notion of validity mentioned here (general, diagonal, and real-world).<sup>35</sup>

Thus, if truths of the form  $\mathbf{A}\varphi$  are taken into consideration, and given the fact that such truths *are* counterfactually necessary,  $\mathbf{N}$  cannot be adequate as an account of “no-choice-about” for reasons that are analogous to the ones undermining  $N$ : if  $\mathbf{A}\varphi$  is true at all, there is no world most similar to the actual one such that, if taken as actual,  $\mathbf{A}\varphi$  would be false, because there simply is no world taken as actual at which  $\mathbf{A}\varphi$  is false.

This problem could seemingly be avoided though by moving up to a three-dimensional semantics in which a new necessity operator, say,  $[3]$ , can be defined as guaranteeing truth at every triple  $w', w', w'$ , and a corresponding subjunctive conditional, say,  $[3]\rightarrow$ , according to which  $\varphi [3]\rightarrow \psi$  would (non-vacuously) hold at a triple  $w, v, z$  just in case at the closest worlds  $w'$  to  $w$ , both  $\varphi$  and  $\psi$  hold at  $w', w', w'$ . More formally:

$$\begin{aligned} \mathfrak{M}, w, v, z \models [3]\varphi & \iff \text{for all } w' \in W, \mathfrak{M}, w', w', w' \models \varphi \\ \mathfrak{M}, w, v, z \models (\varphi [3]\rightarrow \psi) & \iff \text{for all } w' \in W, \text{ if } wR_\varphi w', \text{ then } \mathfrak{M}, w', w', w' \models \psi \end{aligned}$$

Then there could be a new operator for “no-choice-about”, which is in turn defined in terms of  $[3]\rightarrow$ . Yet, the same issue would still arise, for a new actuality operator, say,  $\mathbf{A}$ , defined relative to the triple  $w*, w*, w*$  could be added to the language, too, for reasons that are similar to the ones motivating  $\mathbf{A}$  and  $\mathbf{A}$ , and the need for four dimensions, and so on, would be evident. For this new actuality operator would validate  $\mathbf{A}\varphi \rightarrow [3]\mathbf{A}\varphi$ , while we would want to say that some true instances of  $\mathbf{A}\varphi$  are up to us.<sup>36</sup>

---

<sup>35</sup>Proof: suppose that  $\mathfrak{M}, w, v \models \mathbf{A}\varphi$ , in which case  $\mathfrak{M}, w*, w* \models \varphi$ , by the truth conditions for  $\mathbf{A}$ -formulas. Now assume that  $\mathfrak{M}, w, v \not\models \blacksquare\mathbf{A}\varphi$ , to derive a contradiction. Then  $\mathfrak{M}, w, v \models \neg\blacksquare\mathbf{A}\varphi$ , and so there is a possible world  $v$  in  $\mathfrak{M}$  such that  $\mathfrak{M}, v, v \models \neg\mathbf{A}\varphi$ . But then  $\mathfrak{M}, v, v \not\models \mathbf{A}\varphi$ , whence  $\mathfrak{M}, w*, w* \not\models \varphi$ , by the truth conditions for  $\mathbf{A}$ -formulas, which in turn implies that  $\mathfrak{M}, w*, w* \models \neg\varphi$ , contradicting the assumption. Note: because general implies diagonal and real-world validity, this argument establishes that  $\mathbf{A}\varphi \rightarrow \blacksquare\mathbf{A}\varphi$  is valid according to all three notions of validity mentioned here.

<sup>36</sup>As an anonymous reviewer suggested, one could also add a set of worlds, say,  $D$ , to the class of models, and an operator  $[3]_D$  that diagonalizes only over the points in  $D$ . Then we can have sentences that are necessary in the usual sense even though they are *choice-contingent*, i.e. contingent relative to the points in  $D$ . In this case, there is a sense in which  $D$  could be said to represent the agent’s available (free) options even if they include necessary truths. Although we do not have the space needed to develop this suggestion further, see, for example, the (act-conditional) semantics for ‘can’ offered in Mandelkern et al. (2017), which involves some form of quantification over a contextual set of actions/options. Also, given a contextually supplied set of options for an agent, see Hedden (2012) for a theory of options underlying the subjective *ought*.

---

## 6 Final remarks

The main point of this paper was to raise a problem for the most sophisticated version of the rule BETA-2, formalized by Pruss, and hence to the consequence argument itself. This objection involved actual truths and rigidified descriptions. We believe the cases considered here provide compelling reasons to hold that there are at least some necessary truths that are up to us, the possibility of which is ruled out by BETA-2.

We initially tried to solve this problem by offering a two-dimensional formulation of the Lewisian semantics for subjunctive conditionals or counterfactuals, delivering what we called the counterfactual interpretation of “no-choice-about”. Moving to a two-dimensional semantics for counterfactuals seemed, as least to us, the most promising route to solve the idiosyncratic issues raised by the actuality operator (and, of course, rigidified descriptions). Even though the new semantics was successful in dealing with the counter-examples involving actual truths, we argued that new counter-examples could be generated, and likewise in higher dimensions.

So is this a defeat for the counterfactual sufficiency interpretation? Perhaps. But perhaps one can find reasons to reject the introduction of  $\mathbf{A}$  in the first place, as  $\textcircled{\ast}$  plays an important role in a widespread philosophical literature, whereas  $\mathbf{A}$  does not. Still, there would have to be reasons to reject  $\mathbf{A}$  that are not also reasons to reject its counterpart  $\textcircled{\ast}$  in the one-dimensional case, which might be difficult given their structural similarities in their respective languages. At least for those who think the actuality operator is a coherent and useful tool in philosophical logic, or have no qualms about rigidified descriptions, the present arguments provide compelling reasons for rejecting transfer principles like BETA-2.

## Acknowledgments

We would like to thank two anonymous referees for many helpful comments and suggestions that improved this paper. We are also grateful to Kai Wehmeier for valuable advice. Pedro Merluzzi’s research is funded by São Paulo Research Foundation - FAPESP - (Grant No. 2017/20532-8), with additional support from the LATAM Free Will, Agency and Responsibility project (Grant No. 61255) funded by the John Templeton Foundation.

---

## References

- Berto, F., French, R., Priest, G., Ripley, D. (2018). Williamson on Counterpossibles. *Journal of Philosophical Logic*, 47(4): 693-713.
- Brogaard, B. and Salerno, J. (2007). Why Counterpossibles are Not Trivial. *The Reasoner*, 1: 5-6.
- Brogaard, B. and Salerno, J. (2013). Remarks on Counterpossibles. *Synthese*, 190: 639-60.
- Chalmers, D. (2004). Epistemic Two-Dimensional Semantics. *Philosophical Studies*, 118: 153-226.
- Chalmers, D. (2006). The Foundations of Two-Dimensional Semantics. In M. García-Carpintero and J. Macià (eds.), *Two-Dimensional Semantics*, Oxford: Oxford University Press, pp. 55-140.
- Chellas, B. (1975). Basic Conditional Logic. *Journal of Philosophical Logic*, 4: 133-153.
- Crossley, J. N. and Humberstone, L. (1977). The Logic of “Actually”. *Reports on Mathematical Logic*, 8: 11-29.
- Dasgupta, S. (2014). On the Plurality of Grounds. *Philosophers’ Imprint*, 14: 1-28.
- Davies, M. (1983). Actuality and Context Dependence II. *Analysis*, 43: 128-133.
- Davies, M. (2004). Reference, Contingency, and the Two-Dimensional Framework. *Philosophical Studies*, 118: 83-131.
- Davies, M., and Humberstone, I. L. (1980). Two Notions of Necessity. *Philosophical Studies*, 38(1): 1-30.
- Evans, G. (1979). Reference and Contingency. *The Monist*, 62: 161-189.
- Fara, M. and Williamson, T. (2005). Counterparts and Actuality.

---

*Mind*, 114(453): 1-30.

Finch, A. and Warfield, T. A. (1998). The MIND Argument and Libertarianism. *Mind*, 127: 515-528.

Fritz, P. (2014). What is the Correct Logic of Necessity, Actuality and Apriority? *The Review of Symbolic Logic*, 7(3): 385-414.

Fusco, M. (2019). Naturalizing Deontic Logic: Indeterminacy, Diagonalization, and Self-Affirmation. *Philosophical Perspectives*, 32: 165-187.

Hazen, A. P. (1976). Expressive Completeness in Modal Language. *Journal of Philosophical Logic*, 5: 25-46.

Hedden, B. (2012). Options and the Subjective *Ought*. *Philosophical Studies*, 158: 343-360.

Kaplan, D. (1989). Afterthoughts. In J. Almog, J. Perry, and H. Wettstein (eds.), *Themes from Kaplan*, Oxford: Oxford University Press, pp. 567-614.

Kearns, S. (2011). Responsibility for Necessities. *Philosophical Studies*, 155: 307-24.

Kratzer, A. (1979). Conditional Necessity and Possibility. In R. Bauerle, U. Egli, and A. von Stechow (eds.), *Semantics from Different Points of View*, Heidelberg: Springer, pp. 117-147.

Kripke, S. (1980). *Naming and Necessity*. Cambridge: Harvard University Press.

Lampert, F. (2018). Actuality and the A Priori. *Philosophical Studies*, 175: 809-830.

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

Lewis, D. (1981). Are We Free to Break the Laws? *Theoria*, 47: 113-21.

McKay, T. and Johnson, D. (1996). A Reconsideration of an Argument Against Compatibilism. *Philosophical Topics*, 24: 113-122.

- 
- Mandelkern, M, Schultheis, G, and Boylan, D. (2017). Agentive Modals. *The Philosophical Review*, 126 (3): 301-343.
- Nolan, D. (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, 38: 535-72.
- Nolan, D. (2013). Impossible Worlds. *Philosophy Compass*, 8: 360-72.
- Percival, P. (1991). Knowability, Actuality, and the Metaphysics of Context-Dependence. *Australasian Journal of Philosophy*, 69(1): 82-97.
- Pruss, A. R. (2013). Incompatibilism Proved. *Canadian Journal of Philosophy*, 43: 430-437.
- Restall, G. (2012). A Cut-Free Sequent System for Two-Dimensional Modal Logic, and why it Matters. *Annals of Pure and Applied Logic*, 163: 1611-1623.
- Rosen, G. (2010). Metaphysical Dependence: Grounding and Reduction. In B. Hale and A. Hoffmann (eds.), *Modality*, Oxford, Oxford University Press, pp. 109-135.
- Stalnaker, R. (1968). A Theory of Conditionals. In N. Rescher (ed.), *Studies in Logical Theory*. Oxford: Basil Blackwell.
- Stefánsson, H. O. (2018). Counterfactual Skepticism and Multidimensional Semantics. *Erkenntnis*, 83(5): 875-898.
- Trogon, K. (2013). Grounding: Necessary or Contingent? *Pacific Philosophical Quarterly*, 94: 465-485.
- Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.
- Weatherson, B. (2001). Indicatives and Subjunctives. *Philosophical Quarterly*, 51: 200-216.
- Wehmeier, K. (2004). In the Mood. *Journal of Philosophical Logic*,

---

33: 607-630.

Wehmeier, K. (2013). Subjunctivity and Conditionals. *The Journal of Philosophy*, CX(3): 117-142.

Widerker, D. (1987). On an Argument for Incompatibilism. *Analysis*, 47: 37-41.

Wiggins, D. (1973). Towards a Reasonable Libertarianism. In T. Honderich (ed.), *Essays on Freedom of Action*. Boston: Routledge and Kegan Paul, pp. 43-44.

Witmer, G., Butchard, W., and Trogdon, K. (2005). Intrinsicity Without Naturalness. *Philosophy and Phenomenological Research*, 70: 326-350.

Yablo, S. (2002). Coulda, Woulda, Shoulda. In T. S. Gendler and J. Hawthorne (eds.), *Conceivability and Possibility*. Oxford University Press, pp. 441-492.

Yalcin, S. (2015). Actually, Actually. *Analysis*, 75(2): 185-191.

Zalta, E. (1988). Logical and Analytic Truths that are not Necessary. *Journal of Philosophy*, 85(2): 57-74.