

# Antirealism and the conditional fallacy: a semantic approach\*

Patrick Girard                      Luca Moretti  
p.girard@auckland.ac.nz      l.moretti@abdn.ac.uk

December 21, 2011

## Abstract

The expression conditional fallacy identifies a family of arguments deemed to entail odd and false consequences for notions defined in terms of counterfactuals. The antirealist notion of truth is typically defined in terms of what a rational enquirer or a community of rational enquirers would believe if they were suitably informed. This notion is deemed to entail, via the conditional fallacy, odd and false propositions, for example that there exists necessarily a rational enquirer. If these consequences followed from the antirealist notion of truth, alethic antirealism should probably be rejected. In this paper we analyse the conditional fallacy from a semantic (i.e. model-theoretic) point of view. This allows us to identify with precision the philosophical commitments that ground the validity of this type of arguments. We show that the conditional fallacy arguments against alethic antirealism are valid only if controversial metaphysical assumptions are accepted. We suggest that the antirealist is not committed to the conditional fallacy because she is not committed to some of these assumptions.

## 1 Introduction: the Conditional Fallacy and Antirealism

The expression *conditional fallacy* identifies a family of arguments deemed to entail problematic or lethal consequences for notions analysed in terms of counterfactuals. These arguments apply to analyses of dispositional properties and response-dependence concepts. Saying that an object has a given disposition is roughly saying the object is expected to react in a given way when properly

---

\*We are very grateful to Franz Berto, Fred Kroon, Hannes Leitgeb, Jonathan McKeown-Green, Olivier Roy, Jeremy Seligman, Elia Zardini, audience at the Munich Center for Mathematical Philosophy and a referee of this Journal for valuable discussions and important criticisms upon drafts of this paper. We are grateful to Antti Keskinen and Ralph Wedgwood for drawing our attention to papers relevant for the discussion in Section 6.

stimulated. Counterfactual analyses of dispositions are formulated in terms of a priori (or necessarily) true biconditionals of the form:

Something  $x$  is disposed to give response  $R$  to stimulus  $S$  if and only if  $x$  would give response  $R$  if  $x$  were exposed to stimulus  $S$ .

For instance, a wire is live if and only if electrical current would flow from it to a conductor if the wire were touched by the conductor. The response-dependence analysis states, basically, that the use of certain concepts to describe given objects is licensed by the responses elicited in us by these objects in certain conditions. Response-dependence analyses are typically formulated in terms of a priori (or necessarily) true biconditionals of the form:

Something  $x$  falls under the notion  $N$  if and only if  $x$  would elicit response  $R$  from a subject  $S$  if  $x$  were in conditions  $C$ .

For instance, something is green if and only if it would look green to a standard human subject if it were closely observed by her in normal daylight. Conditional fallacy problems are deemed to emerge, typically, whenever it is true that if  $x$  were placed in the relevant situation  $C$ , this fact would *change*  $x$ 's actual dispositional/eliciting-response properties (cf. Wright [24]: 344-345, Bonevac et al. [2]: 273-282 and Gundersen [7]).<sup>1</sup> In these cases, the conditional analyst would be driven by logically compelling arguments to a false conclusion about  $x$ 's *actual* dispositional or response-dependence features.

Before illuminating this abstract description with an example, let us clarify the basic structure of the alleged logically compelling arguments. The conditional fallacy challenges counterfactual analyses based on the following schema:

$$(SC) \quad \Box(p \equiv (q \Box\rightarrow r))$$

Here  $\Box$  is the necessity operator,  $\equiv$  is the material biconditional, and  $\Box\rightarrow$  is the counterfactual conditional. SC says that, necessarily,  $p$  is the case if and only if were  $q$  to be the case,  $r$  would be the case. Conditional fallacy problems are deemed to emerge whenever  $q$  and  $p$  in SC happen to be related in a way that the truth of  $q$  would affect the actual truth value of  $p$ . The simplest way to account for this relation is to use counterfactuals. So if  $p$  is actually true, a problem arises if (i)  $q \Box\rightarrow \neg p$  (where  $\neg$  is logical negation). And if  $\neg p$  is actually true, a problem arises if (ii)  $q \Box\rightarrow p$ . The contention is that SC in conjunction with either counterfactual produces a valid argument concluding that  $p$  is false when  $p$  is actually true, and *vice versa*. These are the schemata underlying the arguments:<sup>2</sup>

$$(I) \quad \frac{\Box(p \equiv (q \Box\rightarrow r)) \quad q \Box\rightarrow \neg p}{\therefore \neg p} \qquad (II) \quad \frac{\Box(p \equiv (q \Box\rightarrow r)) \quad q \Box\rightarrow p}{\therefore p}$$

<sup>1</sup>The philosophical problem of the conditional fallacy was apparently introduced by Shope [19].

<sup>2</sup>In Lewis-Stalnaker conditional logic, schema I is valid only if  $q$  is possible.

The following case can be interpreted as exemplifying a version of the conditional fallacy based on II. Martin [11] suggests that we might think of defining the property of being live instantiated by a metallic wire through the following instance of SC: necessarily,  $(p)$  a wire  $x$  is live if and only if  $(r)$  electrical current would flow from  $x$  to a conductor  $(q)$  if  $x$  were touched by the conductor. Suppose now that it is the case that  $(\neg p)$   $x$  is a *dead* wire. Imagine however that  $x$  is connected to a reliable machine - an electro-fink - that detects whether  $x$  is touched by a conductor. When such contact occurs, the electro-fink reacts instantaneously by making the wire live for the duration of the contact (this means that  $q \Box \rightarrow p$ ). In this case, II allows us to derive the actually false statement that  $(p)$  the wire  $x$  is live. Johnston [8]’s famous chameleon example is interpretable as a case of conditional fallacy of type I.<sup>3</sup>

The conditional fallacy has been argued to imply lethal consequences for alethic antirealism - shortly, antirealism. In this paper we will primarily focus on these alleged consequences. Since the conditional fallacy objections mounted against antirealism do not depend on I or II, we will not dwell on these two specific schemata here.<sup>4</sup> As we clarify below, the arguments against antirealism depend on *variants* of I and II. In this paper we analyse these variants in depth. As the antirealist disbelieves that truth can be completely evidence-transcendent, she analyses truth in terms of *epistemic* notions such as justification or rational acceptability. The antirealist appears to be committed to the following schema (or a close variant of it):

$$(AR) \quad \Box(T(x) \equiv (Q(x) \Box \rightarrow R(x)))$$

Here  $x$  is a placeholder for declarative statements,  $T(x)$  means ‘it is true that  $x$ ’,  $Q(x)$  means ‘epistemic conditions are suitable for evaluating whether  $x$ ’, and  $R(x)$  means ‘it is rationally believed that  $x$ ’ (or ‘it is justified that  $x$ ’). AR says that, necessarily, it is true that  $x$  if and only if it would rationally be believed that  $x$  if epistemic conditions were suitable for evaluating whether  $x$ . Claiming that epistemic conditions are suitable for evaluating whether  $x$  is claiming - at the very least - that there is a rational enquirer/community suitably positioned or informed for evaluating whether  $x$  is the case. If the antirealist defined truth appealing to *actual* - rather than counterfactual - suitable epistemic conditions, antirealism would have the implausible consequence that very many statements lack truth-value (presumably, all those whose suitable conditions of evaluation do not actually obtain). Since AR has the same propositional structure as SC, antirealism is *prima facie* vulnerable to the conditional fallacy objection.

---

<sup>3</sup>Briefly, a chameleon is in the dark and is actually green. This creature is very shy and very perceptive: if another being were to spot it from a sufficiently close distance, the chameleon would instantaneously notice it and blush bright red. If we suppose that the concept of green is response-dependent and is defined in terms of SC, this is a case in which  $(p)$  the chameleon is actually green but  $(q \Box \rightarrow \neg p)$  if the chameleon were observed, it would not be green any longer. Schema I licenses the false conclusion that  $(\neg p)$  the chameleon is actually not green.

<sup>4</sup>For an overview of the diverse conditional fallacy schemata and a general assessment of them see Bonevac et al. [2].

To our knowledge the first philosopher who has explicitly raised a conditional fallacy objection against antirealism is Wright [24]. Wright focuses on the so-called Peircean conception of truth. On this conception (or family of conceptions) to say that a statement  $x$  is true is to say that  $x$  would rationally be believed if epistemic conditions were *ideal*. Ideal epistemic conditions are, roughly, those in which there exists a rational enquirer/community that has acquired *all evidence* (empirical and rational). On this conception, *any* statement has the same ideal epistemic conditions of evaluation: the end of enquiry. If  $Q$  refers to these epistemic conditions, we can simplify AR into:

$$(AR-P) \quad \Box(T(x) \equiv (Q \Box \rightarrow R(x)))$$

It is *false* that epistemic conditions are actually ideal. We certainly do not possess all evidence (whatever this means). Suppose, however, that  $x$  just says that epistemic conditions are ideal, so  $x = Q$ . As the antirealist is committed to AR-P, the antirealist is also committed to  $\Box(T(Q) \equiv (Q \Box \rightarrow R(Q)))$ . Furthermore, since it is *a priori* true that if  $Q$  (i.e. epistemic conditions are ideal) then  $T(Q)$  (i.e. it is true that epistemic conditions are ideal), then it appears true that  $\Box(Q \supset T(Q))$  (where  $\supset$  is the material conditional), and the antirealist is committed to the latter strict conditional. Wright ([24]: 341-342) has produced a proof reducible to the claim that  $\Box(T(Q) \equiv (Q \Box \rightarrow R(Q)))$  and  $\Box(Q \supset T(Q))$  jointly entail  $T(Q)$ . Since  $Q$  is patently false, this proof apparently blows up antirealism in its Peircean version.

Wright's proof is not completely original but simplifies an earlier demonstration given by Plantinga ([14]: 64-66). Plantinga's proof is reducible to the claim that  $\Box(T(Q) \equiv (Q \Box \rightarrow R(Q)))$  and  $\Box(Q \supset T(Q))$  jointly entail  $\Box T(Q)$ .<sup>5</sup> If this entailment actually holds, the Peircean version of antirealism implies that a *false* statement is not only true, but also *necessarily* true. Plantinga's proof and Wright's proof differ in the strength of their respective modal assumptions: Plantinga apparently obtains a stronger result because he assumes Axiom 4,<sup>6</sup> which is not presupposed by Wright. It is worth stressing, however, that Plantinga is not arguing against antirealism - though he can certainly be reinterpreted to do so. Plantinga takes the conclusion of his proof to show only that the antirealist is committed to some form of *theism* (i.e. to the thesis that, roughly, there is necessarily an omniscient rational enquirer/community). Wright [24] and Brogaard and Salerno [3] respond that this is a misguided interpretation of his finding, which should instead be seen as an instance of the general problem of the conditional fallacy that plagues counterfactual analyses. As we find this response plausible, we will not question it in our paper.

Why does Wright interpret his proof and Plantinga's as instances of the conditional fallacy? Plantinga's and Wright's proofs appear to be licensed by,

<sup>5</sup>Wright's and Plantinga's original proofs rely on exclusively classical principles that are unacceptable for the intuitionist, so an antirealist following Dummett's work could reject them. However, Brogaard and Salerno ([3]) have provide an intuitionistically respectable version of Plantinga's proof, which embeds a version of Wright's proof and is based on the theorem that  $\Box(p \equiv (q \Box \rightarrow r))$  and  $\Box(q \supset p)$  jointly entail  $\Box p$ . See Brogaard and Salerno ([3]: 131)

<sup>6</sup> $\Box p \supset \Box \Box p$ , namely, if  $p$  is necessary then  $p$  is necessarily necessary.

respectively, these inference schemata:

$$(III) \quad \frac{\begin{array}{l} \Box(p \equiv (q \Box \rightarrow r)) \\ \Box(q \supset p) \end{array}}{\therefore p} \quad (IV) \quad \frac{\begin{array}{l} \Box(p \equiv (q \Box \rightarrow r)) \\ \Box(q \supset p) \end{array}}{\therefore \Box p}$$

Both III and IV obtain from inference schema II, considered before, by simply replacing the counterfactual in the second premise with a corresponding strict conditional, and by necessitating the conclusion in case of IV. The similarity of these three inference schemata and of their philosophical uses justify the common expression conditional fallacy to refer to them.

Brogaard and Salerno ([3]: 135-137) have moved a third conditional fallacy objection to antirealism that appears to be grounded on IV.<sup>7</sup> Brogaard and Salerno complain that proofs like Plantinga's and Wright's risk - so to speak - to shoot a dead horse. The point is that these proofs only apply to Peircean versions of antirealism, which demand commitment to the existence of *one* single epistemic situation - the end of enquiry - appropriate for evaluating the truth of *any* statement. Peircean antirealism appears implausible on its own because, to begin with, it is hard to imagine what such a situation would be like.<sup>8</sup> Another problem is that it is even harder to understand how such a situation could be possible for rational enquirers conceived of as humans. If it is acknowledged that the end of enquiry is humanly unattainable, it becomes unclear what explanatory advantage antirealism has over realism. For these reasons, Brogaard and Salerno work out a new conditional fallacy objection that targets more plausible and popular versions of antirealism that allow statements to have *individual* truth-conditions.

Brogaard and Salerno's objection aims to hit directly AR rather than AR-P. Suppose  $p =$  'epistemic conditions are suitable for evaluating whether some statement is true'.<sup>9</sup> Since the antirealist accepts AR, she is committed to  $\Box(T(p) \equiv (Q(p) \Box \rightarrow R(p)))$ . But the antirealist is also committed to  $\Box(Q(p) \supset T(p))$ . Here is a proof. Since  $Q(p)$  means 'epistemic conditions are suitable for evaluating whether  $p$ ', the antirealist is committed to:

Necessarily, if  $Q(p)$  then epistemic conditions are suitable for evaluating whether  $p$ .

Since evaluating whether  $p$  is *a priori* equivalent to evaluating whether it is true that  $p$ , the antirealist ought to accept that:

Necessarily, if  $Q(p)$  then epistemic conditions are suitable for evaluating whether it is true that  $p$ .

Since  $p$  is a statement, the antirealist is committed to:

<sup>7</sup>Brogaard and Salerno's proof improves upon Rea[16]'s proof. Likewise Plantinga, Rea takes his proof to show that the antirealist is committed to some form of theism.

<sup>8</sup>The notion of the end of enquiry and even the notion of approximating to the end of enquiry are not perspicuous and might prove incoherent under close scrutiny. Interestingly, Wright himself ([23]: 44-48) has argued for this conclusion.

<sup>9</sup>An equivalent (and perhaps more precise) formulation of  $p$  is this: 'for some statement  $x$ , epistemic conditions are suitable for evaluating whether  $x$  is true'.

Necessarily, if  $Q(p)$  then epistemic conditions are suitable for evaluating whether some statement is true.

Given that  $p$  means ‘epistemic conditions are suitable for evaluating whether some statement is true’, the antirealist ought to accept that the above strict conditional is equivalent to:

Necessarily, if  $Q(p)$  then  $p$ .

Since ‘ $p$ ’ is *a priori* equivalent to ‘it is true that  $p$ ’, and since  $T(p)$  means ‘it is true that  $p$ ’, the antirealist is committed to inferring that:

Necessarily, if  $Q(p)$  then  $T(p)$ .

In conclusion, the antirealist appears committed to  $\Box(Q(p) \supset T(p))$ . Since the antirealist is committed to both  $\Box(T(p) \equiv (Q(p) \Box \rightarrow R(p)))$  and  $\Box(Q(p) \supset T(p))$ , through IV, the antirealist appears committed to  $\Box T(p)$ . Namely, to the claim that, necessarily, it is true that epistemic conditions are suitable for determining whether some statement is true. These conditions include the existence of a properly placed rational enquirer. The antirealist is thus committed to claiming that, necessarily, it is true that there is a rational enquirer. It is actually true that there are rational enquirers, but this truth appears contingent and not necessary. So it is false that  $\Box T(p)$ .

The efficacy of conditional fallacy objections against antirealism has been questioned in the recent literature. For instance, Moretti [12] has argued that all proofs of the conditional fallacy made against antirealism hinge on the controversial assumption that a counterfactual with impossible antecedent is vacuously true. Moretti has also *conjectured* that any elementary reformulation of these proofs not relying on this assumption will prove incorrect or unsound, or could be dismissed by an antirealist who had reasons to reject a modal system as strong as S5 (or its intuitionistic version). In this paper we criticise the conditional fallacy objection raised against antirealism but we aim to produce a more conclusive result.

A feature common to the proofs cited above is that they all aim to *derive* (syntactically) the conclusion rather than showing that the proofs are (semantically) *valid*. This may encourage the erroneous belief that the conditional fallacy does not depend on metaphysical assumptions but rather follows directly from principles of reasoning which are not peculiar to any specific metaphysical enterprise. Yet once the metaphysical assumptions necessary to III and IV are exposed, the proofs of the conditional fallacy appear questionable. In the following, we focus on no particular proof of III or IV, but on III and IV themselves.<sup>10</sup> We explicitly take a *semantic* approach; this will enable us to identify with accuracy the semantic and thus metaphysical commitments that ground the validity of these inference schemata. We will show that III and IV are valid only under controversial metaphysical assumptions. We will suggest

---

<sup>10</sup>The main reason why we also focus on III, rather than on just IV, is that we cannot exclude a priori that there exists a *false* statement  $s$ , similar to that employed by Brogaard and Salerno, such that  $\Box(Q(s) \supset T(s))$ . If this were the case, one could use III to produce a conditional fallacy argument against, not only Peircean antirealism, but also more plausible forms of antirealism allowing statements to have *individual* truth-conditions. Wright ([24]: 343) would seem to conjecture the existence of statements like  $s$ .

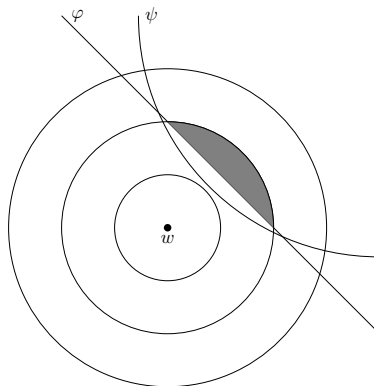


Figure 1: Graphical representation of the similarity operator with the limit assumption as one system of spheres.

that the alethic antirealist who accepts AR (or its more specific variant AR-P) is not committed to these assumptions. Hence, she is not committed to the conditional fallacy.

The structure of the paper is the following: In Section 2 we detail the logical framework that we use in the paper. In Section 3 we introduce semantic (or metaphysical) principles helpful for the analysis of the conditions of validity of III and IV. In Section 4 we single out one condition sufficient for the validity of III given undemanding background assumptions, and one condition sufficient for the validity of IV given the same background assumptions. In Section 5 we show that these two conditions are also respectively necessary for the validity of III and IV given undemanding background assumptions. In Section 6 we suggest metaphysical reasons to doubt that these two conditions are satisfied. In Section 7 we draw the conclusions of the paper.

## 2 Modal Conditional Logic

Call *Modal Conditional Logic* (MCL) any modal logic enriched with logical resources for counterfactuals (or subjunctive conditionals). The language of MCL, which we refer to as  $\mathcal{L}_{\text{MCL}}$ , is based on a set PROP of propositional variables which can be combined recursively using the propositional connectives  $\neg$  and  $\vee$  (where the latter is logical disjunction), and the modalities  $\Box$  and  $\Box \rightarrow$ . The remaining connectives are defined in the usual way. For instance, the *material conditional*  $\varphi \supset \psi$  is defined as  $\neg\varphi \vee \psi$ , logical *conjunction*  $\varphi \wedge \psi$  is defined as  $\neg(\neg\varphi \vee \neg\psi)$ , and the *possibility* operator  $\Diamond\varphi$  by  $\neg\Box\neg\varphi$ .

The most common semantics for MCL is the one introduced by Lewis in [9] and is represented by a family of systems of spheres in which each system is like in Figure 1. Formally, each system of spheres of the family is a partial order

$\leq_w$  defined over the same set of possible worlds  $W$ .<sup>11</sup> We call  $\leq_w$  a similarity relation and read  $u \leq_w v$  as ‘ $u$  is at least as similar to  $w$  as  $v$  is’. Lewis assumes that for each  $w$  in  $W$ , if two worlds  $u$  and  $v$  in  $W$  are comparable in similarity with respect to  $w$  - i.e. if  $u \leq_w v$  - both  $u$  and  $v$  are *accessible* from  $w$ . Using different accessibility relations allows us to discriminate among different types of possibility (or modality). For instance, the worlds that are *physically* possible for  $w$  are accessible from  $w$  through the relation of physical possibility, the worlds that are *metaphysically* possible for  $w$  are accessible from  $w$  through the relation of metaphysical possibility, and so on. As the focus of our discussion is specifically on metaphysical possibility, we will consider only one relation of accessibility  $R$  on  $W$  that identifies metaphysical possibility. Contrary to Lewis, we prefer to stand neutral on whether or not the worlds in  $W$  comparable in similarity with respect to a third world in  $W$  are also (metaphysically) accessible from the latter. This assumption is not mandatory if the accessibility relation is not universal (i.e. as long as it is false that each world in  $W$  has access to *any* world in  $W$ ). As we will see in Section 6, there are reasons for doubting that  $R$  is universal in this sense, and reasons for believing that worlds comparable in similarity with respect to another are not always accessible from the latter.

Some logicians - prominently Stalnaker [20] - impose an assumption on each system of spheres - called *limit assumption* - which states that the strict sub-relation  $<_w$  (which we read ‘... is more similar to  $w$  than ... is’) of  $\leq_w$  is well-founded.<sup>12</sup> The limit assumption guarantees that for any statement  $\varphi$  true at some world of  $\leq_w$ , there is always at least one world  $u$  in  $\leq_w$  with  $\varphi$  true which is *most* similar to  $w$ . In other words,  $u$  is such that there is no world  $v$  in  $\leq_w$  with  $\varphi$  true and such that  $v <_w u$ . We endorse the limit assumption to simplify our discussion, though we believe that our formal results would also obtain without this assumption.<sup>13</sup> In a system of spheres  $\leq_w$ , for any statement  $\varphi$ , the worlds with  $\varphi$  true that are most similar to  $w$  are called the *minimal  $\varphi$ -worlds* or the *closest  $\varphi$ -worlds*. More precisely, we say that a world  $v$  is a minimal  $\varphi$ -world in  $\leq_w$  if and only if  $w \leq_w v$ ,  $\varphi$  is true at  $v$ , and there is no world  $u$  with  $u <_w v$  and such that  $\varphi$  is true at  $u$ . Saying that  $u$  is a *minimal world* in  $\leq_w$  without further specification is saying that  $u$  is a minimal  $\top$ -world in  $\leq_w$ , where  $\top$  is a tautology. A world  $v$  such that  $u \leq_w v$  or  $v \leq_w u$  is said to be *comparable* to  $u$  in  $\leq_w$ ; if neither disjunct is true,  $v$  is said to be *incomparable* to  $u$  in  $\leq_w$ .

According to MCL semantics, a counterfactual is true in a world  $w$  if and only

---

<sup>11</sup>A binary relation  $S$  is a partial order if and only if  $S$  is:

1. Reflexive:  $\forall x : Sxx$ ,
2. Antisymmetric:  $\forall x, y : (Sxy \ \& \ Syx) \Rightarrow x = y$ , and
3. Transitive:  $\forall x, y, z : (Sxy \ \& \ Syz) \Rightarrow Sxz$ .

<sup>12</sup>The strict sub-relation  $<_w$  is defined by:  $u <_w v$  iff  $u \leq_w v$  &  $v \not\leq_w u$ . A binary relation  $S$  is well-founded on a set  $U$  if and only if for every non-empty subset  $U'$  of  $U$ , there exists an element  $x \in U'$  such that, for every  $y \in U'$ , it is false that  $Syx$ .

<sup>13</sup>For example, they would obtain on the minimal semantic definition of  $\Box \rightarrow$  first proposed in Veltman [21] and Burgess [4].

if its consequent is true a all possible worlds closest to  $w$  in which its antecedent is true. (If the antecedent is true at no possible world in  $\leq_w$ , the counterfactual is vacuously true at  $w$ ). In Figure 1, the shaded region represents the minimal  $\varphi$ -worlds. Each of these worlds is also a  $\psi$ -world, so  $\varphi \Box \rightarrow \psi$  is true at  $w$  in this model. On this semantics, a necessitated statement  $\Box\varphi$  is true in  $w$  if and only if  $\varphi$  is true in all worlds accessible from  $w$ . We now make this informal description more precise.

Models for MCL are based on frames  $\mathfrak{F} = \langle W, R, \{\leq_w\}_{w \in W} \rangle$ , where  $W$  is a set of worlds,  $R$  is an *accessibility* relation on  $W$  and  $\{\leq_w\}_{w \in W}$  is a family of *similarity* relations on  $W$ , one for each world in the model. In the remainder of the paper, we call  $R$  the *accessibility relation* and  $\leq$  the *similarity relation*, where  $\leq$  refers to *all* similarity relations in a frame generally considered.

A *model*  $\mathfrak{M}$  is a tuple  $\langle \mathfrak{F}, V \rangle$ , with  $\mathfrak{F}$  a frame and  $V$  a propositional valuation which assigns sets of worlds to each propositional variables of the set  $\text{PROP} - V(p)$  is the set of worlds in which  $p$  is true. We base our interpretation of arbitrary formulae of the language  $\mathcal{L}_{\text{MCL}}$  on this propositional valuation and we define the notion of *satisfaction at a world in a model*, written  $\mathfrak{M}, w \models \varphi$ , recursively:

$$\begin{array}{ll}
\mathfrak{M}, w \models p & \text{iff } w \in V(p) \\
\mathfrak{M}, w \models \neg\varphi & \text{iff } \mathfrak{M}, w \not\models \varphi \\
\mathfrak{M}, w \models \varphi \vee \psi & \text{iff } \mathfrak{M}, w \models \varphi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \Box\varphi & \text{iff } \mathfrak{M}, u \models \varphi \text{ for every } u \text{ such that } Rwu \\
\mathfrak{M}, w \models \varphi \Box \rightarrow \psi & \text{iff } \mathfrak{M}, v \models \psi \text{ for every } \varphi\text{-world } v \text{ minimal in } \leq_w .
\end{array}$$

We say that  $\varphi$  is *true at a world  $w$  in a model  $\mathfrak{M}$* , or *satisfied at  $w$  in  $\mathfrak{M}$*  if and only if  $\mathfrak{M}, w \models \varphi$ . We also say that  $\varphi$  is *valid in a frame  $\mathfrak{F}$*  if and only if  $\varphi$  is satisfied at every world  $w$  in every model  $\mathfrak{M}$  based on  $\mathfrak{F}$ . Finally, we say that  $\varphi$  is *valid in a class of frames  $K$*  if and only if  $\varphi$  is satisfied at every world  $w$  in every model  $\mathfrak{M}$  based on any frame in the class  $K$ .

### 3 Useful Semantic Principles

We would like to introduce here five semantic principles - most of which are probably quite familiar to the reader - that will prove useful for the analysis of the conditions of validity of III and IV carried out in Sections 4 and 5, and for the philosophical discussion of Section 6. These principles are: (1) weak centering of  $\leq$ , (2) reflexivity of  $R$ , (3) transitivity of  $R$ , (4) symmetry of  $R$ , (5) relation containment of  $\leq$  inside  $R$ .

*Weak centering* of  $\leq$  is the assumption that no world in  $W$  can be closer or more similar to a world  $w$  in  $W$  than  $w$  itself. Formally:  $\forall w, v \in W : w \leq_w w \ \& \ (v \leq_w w \Rightarrow w \leq_w v)$ . Thus weak centering says that any world  $w$  is amongst the most similar worlds to  $w$ . This assumption is sometimes dropped in doxastic conditional logic, since a subject might be misguided about the real world and think she lives in a world metaphysically far remote from the real world. However, weak centering is mostly innocuous from a metaphysical point

of view, for it appears true that no matter what a possible world might be, nothing can be more similar to it than itself.<sup>14</sup>

*Reflexivity* of  $R$  is also harmless. It states that each world is accessible to itself. Formally:  $\forall w \in W : Rww$ . Reflexivity of  $R$  is semantically equivalent to Axiom T:  $\Box\varphi \supset \varphi$  (i.e. if  $\varphi$  is necessary, then  $\varphi$  is true), which is customarily assumed in metaphysical contexts because metaphysical necessity is factive.

*Transitivity* of  $R$  states that if a world  $v$  is accessible from a world  $w$ , and a world  $u$  is accessible from  $v$ , then  $u$  is also accessible from  $w$ . Formally:  $\forall w, v, u \in W : (Rwv \& Rvu) \Rightarrow Rwu$ .

*Symmetry* of  $R$  states that if a world  $v$  is accessible from a world  $w$ , then  $w$  is accessible from  $v$ . Formally:  $\forall w, v \in W : Rwv \Rightarrow Rvw$ .

*Relation containment* of  $\leq$  inside  $R$  is the assumption - already considered in the former section - that for every world  $w$ , if two worlds  $v$  and  $u$  are comparable in similarity with respect to  $w$ , then  $v$  and  $u$  are accessible from  $w$ . Formally:  $\forall w, v, u \in W : u \leq_w v \Rightarrow (Rwv \& Rwu)$ . If  $R$  specifically picks up metaphysical modality, this means that any two worlds that can be compared in the similarity relation with respect to  $w$  must be metaphysically possible for  $w$ .

We will see in Section 6 that transitivity of  $R$ , symmetry of  $R$  and relation containment are philosophically more controversial than reflexivity of  $R$  and weak centering. While the antirealist is very probably committed to weak centering and reflexivity of  $R$ , she is presumably not committed to the remaining principles.

## 4 Proving the Conditional Fallacy

Recall that we are interested in these two forms of the conditional fallacy:

$$(III) \quad \frac{\Box(p \equiv (q \Box \rightarrow r)) \quad \Box(q \supset p)}{\therefore p} \quad (IV) \quad \frac{\Box(p \equiv (q \Box \rightarrow r)) \quad \Box(q \supset p)}{\therefore \Box p}$$

In this section, we first show that if weak centering is assumed, there exists one condition - call it *special containment of  $\leq$  in  $R$*  - the satisfaction of which suffices for the validity of III. Then we show that if weak centering is assumed, there exists one condition - call it *quasi transitivity of  $R$  given  $\leq$*  - the satisfaction of which suffices for the validity of IV. In the next section, we will then show that if weak centering and reflexivity of  $R$  are assumed, the satisfaction of special containment is also necessary for the validity of both III and IV, and that the satisfaction of quasi transitivity is also necessary for the validity of IV.

Special containment of  $\leq$  in  $R$  is a special case of relation containment of  $\leq$  in  $R$ ; it says that if a world  $x$  is no less similar to itself than a world  $y$ , then  $y$  is accessible from  $x$ ; formally:

<sup>14</sup>Weak centering should not be confused with the more demanding assumption of *strong centering*, which states that any world is more similar to itself than any other world. Formally:  $\forall w, v \in W : w \leq_w w \& (w \neq v \Rightarrow w <_w v)$ .

$$\forall xy \in W : x \leq_x y \Rightarrow Rxy$$

**Theorem 1** *For any class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering, if special containment holds for every  $F$ , then III is valid in  $K$ .*

Proof: Let us prove this by contradiction. Consider an arbitrary class of frames  $K$  such that each frame has weak centering and special containment. Take an arbitrary  $F \in K$  and an arbitrary model  $M$  based on  $F$  such that for a world  $w \in W$ ,  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$  and  $M, w \models \Box(q \supset p)$ . Finally, assume for the sake of contradiction that  $M, w \not\models p$ . Since  $F$  has weak centering,  $x \leq_x x$  for any  $x \in W$ . So special containment of  $F$  implies that for any  $x \in W$ ,  $Rxx$  - i.e.  $F$  has  $R$  reflexive. Since  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$ ,  $M, w \models \Box(q \supset p)$ , and  $R$  is reflexive,  $M, w \models q \supset p$  and  $M, w \models p \equiv (q \Box \rightarrow r)$ . So, considering that we have assumed  $M, w \not\models p$ , it follows that  $M, w \not\models q$  and  $M, w \not\models q \Box \rightarrow r$ . The latter statement entails that there must be a  $q$ -world  $v$  minimal in  $\leq_w$  such that  $M, v \models q$  and  $M, v \not\models r$ . By the definition of minimality it follows that  $w \leq_w v$ , so special containment of  $F$  implies that  $Rwv$ . Hence, given that  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$  and  $M, w \models \Box(q \supset p)$ , it follows that  $M, v \models p \equiv (q \Box \rightarrow r)$  and  $M, v \models q \supset p$ . Since  $M, v \models q$ , the second consequence entails that  $M, v \models p$ . Let us now derive the negation of this claim. Since  $F$  has weak centering and  $M, v \models q$ ,  $v$  is a minimal  $q$ -world in  $\leq_v$ . We saw before that  $M, v \not\models r$ ; hence  $M, v \not\models q \Box \rightarrow r$ . Since  $M, v \models p \equiv (q \Box \rightarrow r)$ , it follows that  $M, v \not\models p$ , a contradiction. ♣

**Corollary 1** *Any class of frames such that each frame has weak centering and relation containment validates III.*

Proof: This trivially follows from Theorem 1 because special containment is a special case of relation containment. ♣

Let us turn to schema IV. Quasi transitivity of  $R$  given  $\leq$  says that if a world  $y$  is accessible from another  $x$ , and  $y$  is no less similar to itself than a world  $z$ , then  $z$  is accessible from  $x$ . Formally,

$$\forall xyz \in W : (Rxy \ \& \ y \leq_y z) \Rightarrow Rxz$$

We call this condition quasi transitivity because if the second conjunct of its antecedent, i.e.  $y \leq_y z$ , is replaced by  $Ryz$  we obtain transitivity of  $R$ .

**Theorem 2** *For any class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering, if quasi transitivity holds for every  $F$ , then IV is valid in  $K$ .*

Proof: Let us prove this by contradiction. Consider an arbitrary class of frames  $K$  such that each frame in  $K$  has weak centering. Assume also that quasi transitivity holds for every frame in  $K$ . Take an arbitrary frame  $F \in K$  and an arbitrary model  $M$  based on  $F$  such that for a world  $w \in W$ ,  $M, w \models$

$\Box(p \equiv (q \Box \rightarrow r))$  and  $M, w \models \Box(q \supset p)$ . Assume for the sake of contradiction that  $M, w \not\models \Box p$ . It follows from the last three assumptions that there is a world  $v \in W$  such that  $Rwv$  and  $M, v \not\models p, M, v \models p \equiv (q \Box \rightarrow r)$  and  $M, v \models q \supset p$ . Hence,  $M, v \not\models q \Box \rightarrow r$  and  $M, v \not\models q$ . Given that  $M, v \not\models q \Box \rightarrow r$ , there is a  $q$ -world  $u$  minimal in  $\leq_v$  such that  $M, u \models q$  and  $M, u \not\models r$ . Note also that since  $Rwv$  and  $v \leq_v u$ , given that quasi transitivity holds in  $F$ , we obtain that  $Rwu$ . Thus, since  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$  and  $M, w \models \Box(q \supset p)$ , it follows that  $M, u \models p \equiv (q \Box \rightarrow r)$  and  $M, u \models q \supset p$ . From the latter claim, as  $M, u \models q$ , it follows that  $M, u \models p$ . Let us now derive the negation of this claim. By weak centering of  $F$ , since  $M, u \models q$ ,  $u$  is a minimal  $q$ -world in  $\leq_u$ . We proved before that  $M, u \not\models r$ ; thus  $M, u \not\models q \Box \rightarrow r$ . Given that  $M, u \models p \equiv (q \Box \rightarrow r)$ , it follows that  $M, v \not\models p$ , a contradiction. ♣

Theorem 2 enables us to determine a condition sufficient for IV's validity formulated in terms of weak centering, relation containment and transitivity of  $R$ .

**Corollary 2** *Any class of frames such that each frame has weak centering, relation containment and transitivity of  $R$  validates IV.*

Proof: Consider an arbitrary class  $K$  of frames such that each frame has weak centering, relation containment and transitivity of  $R$ . Take an arbitrary frame  $F \in K$  and consider three arbitrary worlds  $u, v, w \in W$  such that  $Ruv$  and  $v \leq_v w$ . Relation containment and  $v \leq_v w$  jointly imply that  $Rvw$ . Since transitivity of  $R$  holds in  $F$ ,  $Ruv$  and  $Rvw$  jointly entail that  $Ruw$ . Thus any arbitrary  $F$  in  $K$  has quasi transitivity. Thus, any class  $K$  of frames such that each frame has weak centering, relation containment and transitivity of  $R$  satisfies quasi transitivity for any frame. By Theorem 2, any such class validates IV. ♣

## 5 Defusing the Conditional Fallacy

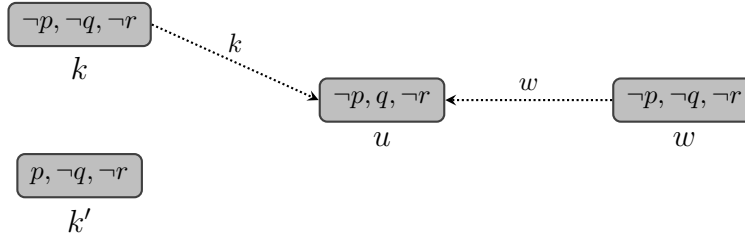
Model-theory can be used to understand to what extent special containment and quasi transitivity are crucial in validating the proofs of the conditional fallacy against antirealism. We will now show that, given very weak background assumptions, dropping special containment *invalidates* both III and IV, and dropping quasi transitivity *invalidates* IV.

**Theorem 3** *For any class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering and reflexivity of  $R$ ,  $K$  validates III only if special containment holds for every  $F$ .*

Proof: We show this by proving that in an arbitrary class of frames  $K$  such that each of its frames has weak centering and reflexivity of  $R$ , if special containment is not satisfied by an arbitrary frame  $F$  in  $K$ , III is not valid in

$F$ . Suppose then that  $F$  is an arbitrary frame.  $F$  must have the following features:  $\leq$  satisfies weak centering,  $R$  is reflexive, and there exist two worlds  $w, u \in W$  such that it is true that  $w \leq_w u$  but false that  $Rwu$ . The strategy is to build a model on top of this  $F$  by giving a propositional valuation that satisfies the premises of III but not the conclusion, and which is not precluded by any possible additional feature of  $F$ . We consider only the case in which  $w \neq u$  because, since  $R$  is reflexive by assumption,  $Rwu$  necessarily holds if  $w = u$ .

Take a valuation  $V$  such that  $V(p) = \{x \mid x \text{ is incomparable to } u \text{ in } \leq_x\}$  and  $V(q) = \{u\}$ .<sup>15</sup> This is a graphic representation of the model:



The labelled dotted arrow indicates similarity relation - the label stands for the subscripts of  $\leq$ , so the dotted arrow from  $w$  to  $u$  indicates that  $w \leq_w u$ , and the dotted arrow from  $k$  to  $u$  indicates that  $k \leq_k u$ . Here  $k$  stands for any world in  $W$  different from  $u$  but comparable to  $u$  in  $\leq_k$ , and  $k'$  stands for any world in  $W$  incomparable to  $u$  in  $\leq_{k'}$ . No dotted arrow goes from  $k'$  to  $u$ , which reflects the assumption that it is false that  $k' \leq_{k'} u$ . No black arrow - meaning the accessibility relation - goes from  $w$  to  $u$ , which reflects the assumption that it is false that  $Rwu$ . Since weak centering holds in this model, the similarity relations are all reflexive; the accessibility relation is also reflexive, but all reflexive arrows are left implicit and not graphically represented. It is assumed that any world can be linked to any other through the accessibility relation and similarity relations, provided that  $Rwu$  remains false and the distinction between worlds comparable and incomparable to  $u$  is respected.

Note that for every  $x \in W$ ,  $M, x \models q \supset p$  with the sole exception for  $x = u$ . But it is false that  $Rwu$  by assumption. So  $M, w \models \Box(q \supset p)$ . Let us now prove that  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$  by showing that for any  $x \in W$  such that  $x \neq u$ ,  $M, x \models p \equiv (q \Box \rightarrow r)$ . Let us first show that  $M, w \models p \equiv (q \Box \rightarrow r)$ . Note that  $u$  is a minimal  $q$ -world in  $\leq_w$  because, by assumption,  $w \leq_w u$  and  $M, u \models q$ ; furthermore, given that at every world  $x$  different from  $u$  in  $W$ , by assumption  $M, x \not\models q$ , there cannot be any world  $x$  in  $W$  with  $q$  true such that  $x <_w u$ . Since  $u$  is a minimal  $q$ -world in  $\leq_w$  and  $M, u \not\models r$ , then  $M, w \not\models q \Box \rightarrow r$ . As we have assumed  $M, w \not\models p$ , it follows that  $M, w \models p \equiv (q \Box \rightarrow r)$ .

Let us now prove that this biconditional is true at any  $x \in W$  such that  $x \neq w$  and  $x \neq u$ . Consider first any world  $k$  comparable to  $u$  in  $\leq_k$ . In this

<sup>15</sup>That is to say,  $V$  makes  $p$  true at any world  $x$  in  $W$  such that neither  $u \leq_x x$  nor  $x \leq_x u$ , and also  $q$  true at  $u$ .

case,  $k \leq_k u$  or  $u \leq_k k$ . Given that weak centering holds for  $F$ , the second disjunct entails the first. Thus  $k \leq_k u$ . Note also that we have assumed that  $M, u \models q$ . Furthermore, given that at every world  $x$  different from  $u$  in  $W$ , by assumption  $M, x \not\models q$ , there cannot be any world  $x$  in  $W$  with  $q$  true such that  $x <_k u$ . So  $u$  is a minimal  $q$ -world in  $\leq_k$ . From this and the assumption that  $M, u \not\models r$ , we obtain  $M, k \not\models q \Box \rightarrow r$ . Since we have assumed that  $M, k \not\models p$ , it follows that  $M, k \models p \equiv (q \Box \rightarrow r)$ . Consider now any world  $k'$  incomparable to  $u$  in  $\leq_{k'}$ . As  $k'$  is incomparable to  $u$ , it is false that  $k' \leq_{k'} u$ . Since  $u$  is the only world in  $W$  with  $M, u \models q$ , there exists no minimal  $q$ -world to  $k'$ , with the consequence that  $q \Box \rightarrow r$  is vacuously true at  $k'$ . As  $p$  is also true at  $k'$ ,  $M, k' \models p \equiv (q \Box \rightarrow r)$ . Since all worlds in  $W$  verify this biconditional with the sole exception of  $u$ , which is by assumption inaccessible from  $w$ , we can conclude that  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$ . We also saw that  $M, w \models \Box(q \supset p)$  and  $M, w \not\models p$ . Hence III is invalid in  $F$ .  $\clubsuit$

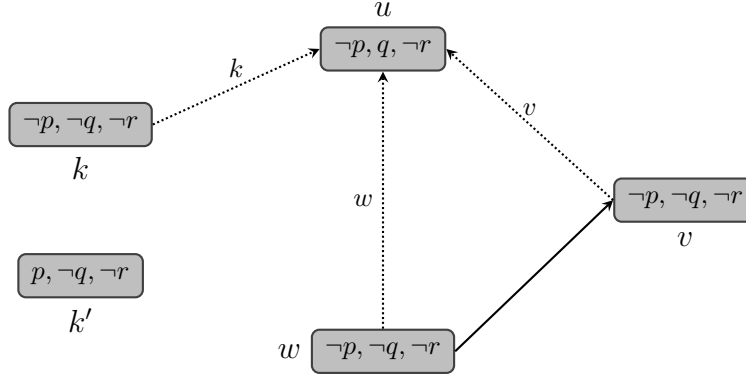
**Theorem 4** *For any class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering and  $R$  reflexive,  $K$  validates IV only if quasi transitivity holds for every  $F$ .*

Proof: We show this by proving that in an arbitrary class of frames  $K$  such that each of its frames has weak centering and  $R$  reflexive, if quasi transitivity is not satisfied by an arbitrary frame  $F$  in  $K$ , IV is not valid in  $F$ . Assume therefore that  $F$  has weak centering and  $R$  reflexive, and that there exist three worlds  $w, v$  and  $u$  in  $W$  such that  $Rwv, v \leq_v u$  but not  $Rwu$ . We use the same strategy as before: we build a model on top of this  $F$  by giving a propositional valuation that satisfies the premises of IV but not its conclusion, and which is not precluded by any possible additional feature of  $F$ . It is easy to verify that, as  $R$  is reflexive,  $Rwu$  can prove false only if (1)  $w = v \neq u$  or (2)  $w \neq v \neq u$ .

Let us first consider case 1. If  $F$  is such that  $w = v \neq u$ , we are back to the situation considered in Theorem 3. That is,  $K$  has weak centering and  $R$  reflexive, and there exist two worlds  $w, u \in W$  such that it is true that  $w \leq_w u$  but not that  $Rwu$ . Take again the propositional valuation used in Theorem 3 that makes both  $\Box(p \equiv (q \Box \rightarrow r))$  and  $\Box(q \supset p)$  true and  $w$  and  $p$  false at  $w$ . As  $Rww, M, w \not\models \Box p$ . So IV is invalid in  $F$ .

Let us turn to case 2. Quasi transitivity says that  $\forall wvu \in W : (Rwv \ \& \ v \leq_v u) \Rightarrow Rwu$ . This principle does not specify whether or not  $w \leq_w u$ . We can thus distinguish between two sub-cases: either (2.1)  $F$  is such that  $w \leq_w u$  or (2.2)  $F$  is such that not  $w \leq_w u$ .

Consider first 2.1. Take again the valuation  $V$  used for Theorem 3.  $V(p) = \{x \mid x \text{ is incomparable to } u \text{ in } \leq_x\}$  and  $V(q) = \{u\}$ . This is a graphic representation of the model:

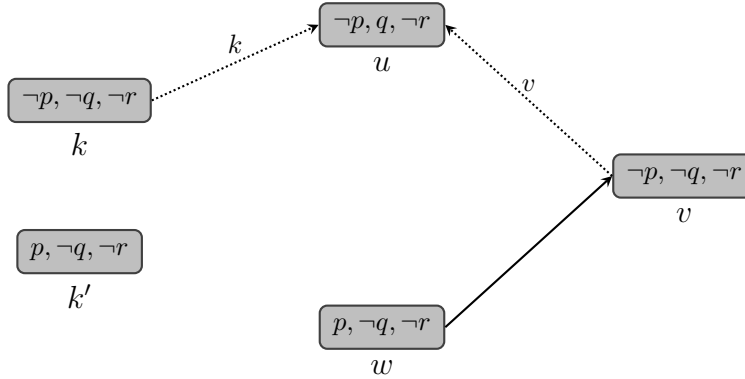


The black arrow from  $w$  to  $v$  indicates that  $Rwv$ . No black arrow goes from  $w$  to  $u$ , which reflects the assumption that it is false that  $Rwu$ . The labelled dotted arrows from  $v$  to  $u$ , from  $w$  to  $u$  and from  $k$  to  $u$  indicate, respectively, that  $v \leq_v u$ ,  $w \leq_w u$  and  $k \leq_k u$ . Here  $k$  stands for any world in  $W$  different from  $w$  and  $u$  comparable to  $u$  in  $\leq_k$ , and  $k'$  stands for any world in  $W$  incomparable to  $u$  in  $\leq_{k'}$ . No dotted arrow goes from  $k'$  to  $u$ , which reflects the assumption that it is false that  $k' \leq_{k'} u$ . The accessibility relation and the similarity relations are assumed to be all reflexive, though this is not graphically represented. It is assumed that any world can be linked to any other through the accessibility relation and similarity relations, provided that  $Rwu$  remains false and the distinction between worlds comparable and incomparable to  $u$  is respected.

Note that for every  $x \in W$ ,  $M, x \models q \supset p$  with the sole exception for  $x = u$ . But it is false that  $Rwu$  by assumption. So  $M, w \models \Box(q \supset p)$ . Let us now prove that  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$  by showing that for any  $x \in W$  such that  $x \neq u$ ,  $M, x \models p \equiv (q \Box \rightarrow r)$ . Let us first show that  $M, w \models p \equiv (q \Box \rightarrow r)$ . Note that  $u$  is a minimal  $q$ -world in  $\leq_w$  because, by assumption,  $w \leq_w u$  and  $M, u \models q$ ; furthermore, given that at every world  $x$  different from  $u$  in  $W$ , by assumption  $M, x \not\models q$ , there cannot be any world  $x$  in  $W$  with  $q$  true such that  $x <_w u$ . Since  $u$  is a minimal  $q$ -world in  $\leq_w$  and by assumption  $M, u \not\models r$ , then  $M, w \not\models q \Box \rightarrow r$ . As we have assumed  $M, w \not\models p$ , it follows that  $M, w \models p \equiv (q \Box \rightarrow r)$ . Let us now prove that this biconditional is true at any  $x \in W$  such that  $x \neq w$  and  $x \neq u$ . Consider first any world  $k$  comparable to  $u$  in  $\leq_k$  - we include  $v$  among these worlds. In this case,  $k \leq_k u$  or  $u \leq_k k$ . Given that weak centering holds for  $F$ , the second disjunct entails the first. Thus  $k \leq_k u$ . Considering that  $M, u \models q$  and that at every world  $x$  different from  $u$  in  $W$ ,  $M, x \not\models q$ , there is no world  $x$  in  $W$  with  $q$  true such that  $x <_k u$ . So  $u$  is a minimal  $q$ -world in  $\leq_k$ . Since  $M, u \not\models r$ , we obtain  $M, k \not\models q \Box \rightarrow r$ . Since we have assumed that for any  $k$ ,  $M, k \not\models p$ , it follows that for any  $k$ ,  $M, k \models p \equiv (q \Box \rightarrow r)$ . It also follows, as a particular case, that  $M, v \models p \equiv (q \Box \rightarrow r)$ . Take finally any world  $k'$  incomparable to  $u$  in  $\leq_{k'}$ . As  $k'$  is incomparable to  $u$ , it is false that  $k' \leq_{k'} u$ . Since  $u$  is the only world in  $W$  with  $M, u \models q$ , there exists no minimal  $q$ -world in  $\leq_{k'}$ , with the consequence that  $q \Box \rightarrow r$  is vacuously true at any  $k'$ . As  $p$  is also

true at  $k'$ ,  $M, k' \models p \equiv (q \Box \rightarrow r)$ . As all worlds in  $W$  verify this biconditional with the sole exception of  $u$ , which is inaccessible from  $w$ , we can conclude that  $M, w \models \Box(p \equiv (q \Box \rightarrow r))$ . We saw that  $M, w \models \Box(q \supset p)$ . Furthermore, by assumption  $Rwv$  and  $M, v \not\models p$ , so that  $M, w \not\models \Box p$ . Hence IV is invalid in  $F$ .

Consider now sub-case 2.2, in which  $F$  does not have  $w \leq_w u$ . Take again the valuation  $V$  used before, such that  $V(p) = \{x \mid x \text{ is incomparable to } u \text{ in } \leq_x\}$  and  $V(q) = \{u\}$ . This is a graphic representation of the model:



The only differences from before are that, now, no dotted arrow links  $w$  to  $u$  - as it is false that  $w \leq_w u$  - and that  $p$  is true at  $w$ . Here  $k$  stands for any world in  $W$  distinct from  $u$  comparable to  $u$  in  $\leq_k$ , and  $k'$  stands for any world in  $W$  incomparable to  $u$  in  $\leq_{k'}$ . The proof that IV is invalid in  $F$  when  $F$  instantiates this sub-case is the same as the proof that IV is invalid in  $F$  when  $F$  instantiates the former sub-case; the only difference is in the demonstration that  $M, w \models p \equiv (q \Box \rightarrow r)$ . To show this now just count  $w$  as one of the  $k'$ -worlds. This complete the proof of Theorem 3. ♣

Here is an interesting corollary that exposes the centrality of special containment in the proofs of the conditional fallacy against antirealism.

**Corollary 3** *For any class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering and  $R$  reflexive,  $K$  validates IV only if special containment holds for every  $F$ .*

Proof: Let us first show that any frame that has  $R$  reflexive and quasi transitivity has also special containment. Suppose  $F$  has  $R$  reflexive and quasi transitivity, and consider two worlds  $v, u \in W$  such that  $v \leq_v u$ . (If there are no such worlds, special containment is vacuously satisfied). Let us stipulate that  $v = w$ . Since  $Rvv$ , then  $Rwv$ . The latter,  $v \leq_v u$  and quasi transitivity jointly imply that  $Rwu$ , that is  $Rvu$ . Therefore,  $F$  has special containment. Consider now a class of frames  $K$  such that each frame  $F$  in  $K$  has weak centering and  $R$  reflexive. Given Theorem 4,  $K$  validates IV only if quasi transitivity holds for

every  $F$ . Considering what just shown, this entails that  $K$  validates IV only if special containment holds for every  $F$ . ♣

To summarise, it turns out that satisfying special containment is a condition necessary for validating both III and IV (given reflexivity of  $R$  and weak centering). On the other hand, since fulfilling special containment (together reflexivity of  $R$  and weak centering) does not imply fulfilling quasi transitivity,<sup>16</sup> satisfying quasi transitivity is an independent necessary condition for validating IV.

What philosophical significance do the above formal results have? It is hard to imagine how the antirealist could drop weak centering or reflexivity of  $R$ . Consequently, the antirealist would commit a conditional fallacy, and would thus run afoul a reductio, *only if* she also accepted special containment, necessary to validate both III and IV. Furthermore, the non-Peircean antirealist would commit a conditional fallacy *only if* she accepted both special containment and quasi transitivity, necessary to validate IV. Now consider for instance the latter principle. Why should the antirealist be committed to maintaining that if a world  $y$  is accessible from another world  $x$  and  $y$  is no less similar to itself than a possible world  $z$ , then  $z$  is accessible from  $x$ ? Whoever makes a conditional fallacy argument against the antirealist should answer this question. It is apparent that the answer - if there is a definite answer - should be metaphysical in nature and not just logical.

Before leaving the logical ground to enter the metaphysical territory, we would like to dispel a possible objection. We have formulated our arguments in a classical setting, whereas the debate on antirealism has been intertwined with the debate between classical and intuitionistic logic. But we need not be concerned with such issues here, as Theorems 3 and 4 all show the invalidity of an argument form in a classical context, from which the invalidity of the same argument form in an intuitionistic setting immediately follows. It would be interesting to formulate an intuitionistic modal logic expanded with counterfactuals and explore how our counter-examples could be adapted to such system, but we leave this (tricky!) task to the reader, as this would distract us from our main topic.

## 6 Metaphysical Reasons to Doubt Special Containment and Quasi Transitivity

The special containment condition and the quasi transitivity condition are expressed by universal conditionals the consequents of which state that a possible world of a given type has access to a possible world. Either conditional can be falsified only if its consequent is falsified. This can happen only if the accessibility relation  $R$  is not universal - i.e. only if it is false that any possible

<sup>16</sup>Consider for instance the frame  $F$  with  $W = \{w, v, u\}$ , weak centering,  $R$  reflexive and such that  $Rwv$ ,  $v \leq_v u$  and  $Rvu$  but not  $Rwu$ .  $F$  satisfies special containment but not quasi transitivity.

world has access to any possible world. In contemporary philosophical logic there is no agreement on whether  $R$  is universal. In the following, we review arguments adduced by philosophers to deny that  $R$  is universal, and we suggest ways in which the antirealist could further develop these arguments to make a case specifically against special containment or quasi transitivity.

Armstrong [1], while defending his *combinatorialist* theory of possibility, has found a reason to reject symmetry of  $R$ ; Nathan Salmon [17, 18] has raised an independent Sorites argument against transitivity of  $R$ . Let us start with Armstrong's argument. A combinatorialist theory of possibility is a metaphysical picture according to which possible worlds are rearrangements or recombinations of elementary individuals and elementary properties existing in the actual world. Simple individuals are those that lack proper parts, and simple properties are those that do not have other properties as constituents. Simple individuals and simple properties exist only contingently and will have to be determined on the grounds of total science. (See Divers [5]: 174-176 and 207-208). Consider a simple property  $P$  actually instantiated by some individual and an actual simple individual  $a$  that does not instantiate  $P$ . The state of affairs that  $Pa$  can be obtained by recombining  $a$  and  $P$ . Thus in the actual world it is *possible* that  $Pa$  - namely, possible worlds in which it is true that  $Pa$  are *accessible* from the actual world. Consider now all re-combinations of the actually instantiated properties with all existing simple individuals. One recombination is to the effect that  $a$  does not exist any more; another recombination is to the effect that  $P$  is no longer instantiated by any individual. These constitute genuine possibilities for the combinatorialist - something that could have happened. Possible worlds in which  $a$  does not exist or  $P$  is not instantiated are thus accessible from the actual world. Consider now a possible world  $v$  that does not include  $a$  or  $P$ . The actual world is not accessible from  $v$  because it cannot be obtained by re-arranging the contingent elements constituting  $v$ , *which do not include a or P*. Hence, for the combinatorialist, the accessibility relation  $R$  is not symmetric and thus not universal.

Combinatorialism is affected by difficulties, but this is true of any interesting theory of possibility. We are not defending combinatorialism. Our point is that the antirealist appears *prima facie* entitled to endorse some version of combinatorialism - for instance, one according to which the contingent recombining items are *mind- or language-dependent*. The antirealist endorsing combinatorialism could adduce Armstrong's argument to contend that  $R$  is not universal.

There are other interesting arguments against the symmetry of  $R$ ; see for instance, among others, Wedgwood [22], Peacocke [13], Dummett ([6]: 328-348) and Quinn [15]. Some of these arguments are less general than Armstrong's, as they presuppose specific views in particular areas of philosophy. For example, Wedgwood ([22]) has made a case that non-reductive physicalism about mental properties - a view quite fashionable today - cannot be true if  $R$  is symmetric. We tend to believe that even "local" arguments of this type can shed doubts on the symmetry and so universality of  $R$ . Versions of these arguments might be adduced by the antirealist.

Let us turn to Salmon's argument. Consider an artefact - for example a

table, which we will call T. It is intuitively plausible that T, while retaining its numerical identity, could have originated from a piece of tree trunk  $W_1$  very slightly different from the piece of tree trunk  $W_0$  from which T has actually originated. Suppose  $W_1$  has the same shape and size as  $W_0$  but is taken from one millimetre further down the same trunk as  $W_0$ . In short, it is *possible* that T is made of  $W_1$ . Consider now that if T had actually originated from  $W_1$ , it would plausibly be true that T could have originated from  $W_2$  - a piece of wood taken from one additional millimetre further down the same trunk. Thus it is possible that it is possible that T is made of  $W_2$ . Let us reiterate this reasoning by one thousand times to reach the apparently correct conclusion that it is possible that it is possible that it is possible ... that T is made of  $W_{1000}$  - a piece of wood that differs from T's actual piece of wood by one meter. If  $R$  is transitive, we can reduce this very long modal claim to the short claim that it is just possible that T is made of  $W_{1000}$ ; we can do so through the reiterated application of the principle of transitivity. But the short claim appears blatantly false: if T had originated from a piece of wood that differs from T's actual piece of wood by one meter, T would be a distinct individual! So we had better not to assume transitivity of  $R$ . Clearly, if  $R$  is not transitive,  $R$  is not universal. We are not endorsing Salmon's argument. Our point is simply that there is apparently no reason why the antirealist could not appeal to Salmon's argument when threatened by the conditional fallacy objection.

As we have said, the antirealist could develop Armstrong's or Salmon's argument to make a case against special containment or quasi transitivity. The antirealist endorsing combinatorialism could for instance argue as follows: suppose a possible world  $u$  is obtained by recombining the actual world  $w$ . Imagine  $u$  differs from  $w$  because many elementary individuals existing in  $w$  do not exist in  $u$  any more, or because many elementary properties instantiated in  $w$  are no longer instantiated in  $u$ .<sup>17</sup> Suppose furthermore that a world  $v$  is obtained through a very minimal recombination of  $u$  - e.g. the only difference between  $v$  and  $u$  is that an elementary individual that instantiates a given elementary property at  $u$  no longer instantiates that property at  $v$ . Given the composition of  $u$ ,  $v$  and  $w$ , it is intuitively correct to judge that  $v$  is *more similar* to  $u$  than  $w$  is. Note that whether or not  $w$  is possible for  $u$ , this appears irrelevant for the intuitive correctness of our judgement. Since  $\leq_u$  is meant to be a similarity relation,  $\leq_u$  should comply with our intuitive judgements of similarity. Thus we should conclude that that  $v \leq_u w$ . Let us now replace  $v$  with  $u$  itself. Since  $u$  and  $v$  are almost identical whereas  $v$  is very much dissimilar from  $w$ , it is correct to conclude that  $u$  is *more similar* to  $u$  itself than  $w$  is, so that  $u \leq_u w$ . Again, whether or not  $w$  is possible for  $u$ , this appears irrelevant for the intuitive correctness of our judgement. As it is false that  $Ruw$ , special containment should be dropped.

The antirealist appealing to combinatorialism could alternatively argue along these lines: suppose  $u$  is obtained by recombination from our actual world  $w$

---

<sup>17</sup>In this and in all other examples of this section we will implicitly presuppose that weak centering is satisfied and that  $R$  is reflexive.

by just eliminating one of the elementary individuals of  $w$ . Since  $u$  is almost identical to our actual world, it is hard to deny that the first of the following counterfactuals appears true at  $u$  while the second appears false at  $u$ :

- (A) If an additional elementary individual existed, no pig would fly.
- (B) If an additional elementary individual existed, some pig would fly.

Although we know that none of the worlds that make the antecedents of A and B true is possible for  $u$ , this fact appears irrelevant when we judge the truth-values of A and B at  $u$ .<sup>18</sup> Modal Conditional Logic (MCL) can match our intuitive judgements of truth-value of A and B at  $u$  only if we assume that (1) there is some world  $x$  such that  $u \leq_u x$  at which it is true that an additional elementary individual exists, and (2) at any world  $x$  of this type that counts as minimal it is also true that no pig fly. If we insist that  $\leq_u$  is contained in  $R$ , so that there exists no world  $x$  of this type, we should conclude that A and B are *both* (vacuously) true, which seems incorrect. As a semantic theory that accounts for our pre-theoretical judgements about the truth-values of counterfactuals like A and B would be *ceteris paribus* preferable to a semantic theory unable to do so, it seems reasonable that MCL should comply with (1) and (2). Since there is some  $x$  such that  $u \leq_u x$  but not  $Rux$ , special containment cannot be accepted.

The antirealist who did not want to commit herself to combinatorialism could probably run a parallel argument that build up on Salmon's case against transitivity of  $R$ . Consider again the table T that has been assumed to have originated from the piece of tree trunk  $W_0$  in our actual world  $w_0$ . It is hard to deny - the antirealist could contend - that the first of the following counterfactuals appears true at  $w_0$  while the second appears false at  $w_0$ :

- (C) If T had originated from  $W_{1000}$ , no pig would fly.
- (D) If T had originated from  $W_{1000}$ , some pig would fly.

Again, knowing that none of the worlds that make the antecedents of these counterfactuals true is possible for  $w_0$  appears irrelevant when we judge the truth-values of C and D. MCL can account formally for these intuitive judgements only if we assume that (1) there is some world  $x$  such that  $w_0 \leq_{w_0} x$  at which it is true that T has originated from  $W_{1000}$ , and (2) at any world  $x$  of this type that counts as minimal it is also true that no pig flies. Since there is some  $x$  such that  $w_0 \leq_{w_0} x$  but not  $Rw_0x$ , special containment should be rejected.

Given that quasi transitivity entails special containment (once weak center-

---

<sup>18</sup>This argument and the next one parallel a well-known objection to Lewis'[9]'s thesis that all counterfactuals with impossible antecedents are vacuously true. For instance, Mares [10] lists counterfactuals with impossible antecedents that appear true and counterfactuals with impossible antecedents that appear false. A difference between Mares' objection and the arguments we put forward here (on the antirealist's behalf) is that ours are meant to apply to counterfactual with *locally* impossible antecedents - i.e. antecedents that are false at any world *accessible from a given possible world* - whereas Mares' objection is meant to apply to counterfactual with *absolutely* impossible antecedents - i.e. antecedents that are false at *any possible world whatsoever*.

ing and reflexivity of  $R$  are assumed), all arguments against special containment considered so far are also arguments against quasi transitivity. To finish, let us consider an alternative way in which the antirealist could exploit Salmon's argument to make a case against quasi transitivity that does not affect special containment - in other words, if this argument against quasi transitivity is successful, special containment need not be dropped. Let us simplify our language: consider again table T originated from piece of three trunk  $W_0$  in the actual world  $w_0$ . For any real  $n$ ,  $w_n$  will hereafter stand for a world in which T has originated from piece of trunk tree  $W_n$  taken from  $n$  millimetres further down the same trunk as  $W_0$ . Consequently, instead of saying that at  $w_n$  it appears possible that T is made of  $W_{n+1}$ , we will simply say that  $w_{n+1}$  appears possible for  $w_n$ . The antirealist could argue as follows:  $w_1$  appears possible for  $w_0$ ;  $w_2$  appears possible for  $w_1$ ;  $w_3$  appears possible for  $w_2$ ; and so on. Salmon shows that if we repeat this pattern by one thousand times and then apply the principle of transitivity by the same number of times, we reach the incorrect conclusion that  $w_{1000}$  appears possible for  $w_0$ . But it is easy to see that we can reach the same conclusion by the reiterated application of the principle of quasi transitivity alone, even if  $R$  is not transitive. Given that  $w_1$  appears possible for  $w_0$ ,  $Rw_0w_1$ . Consider now that  $w_1$  is certainly no less similar to  $w_1$  than  $w_2$  is. So  $w_1 \leq_{w_1} w_2$ . (This claim is not particularly controversial in this case, as  $w_2$  is accessible from  $w_1$ ). Given that  $Rw_0w_1$  and that  $w_1 \leq_{w_1} w_2$ , by quasi transitivity, we infer that  $Rw_0w_2$ . Now reiterate the reasoning:  $w_2$  is certainly no less similar to  $w_2$  than  $w_3$  is. So  $w_2 \leq_{w_2} w_3$ . Given this and the fact that  $Rw_0w_2$ , we obtain via quasi transitivity that  $Rw_0w_3$ . And so on. If we reiterate this reasoning by a sufficient number of times, we will arrive to the intuitively false conclusion that  $Rw_0w_{1000}$ . This is why quasi transitivity cannot be accepted.

To prevent misunderstandings, let us emphasise one more time that we are not *making* or *endorsing* any of the arguments presented in this section. Though we are not taking position on the soundness of these arguments, it seems to us that none of them is obviously fallacious or unsuccessful. Our point is simply that there is apparently no reason why the antirealist could not appeal to some of these arguments, or variants of them, when threatened by the conditional fallacy objection.

## 7 Conclusions

In this paper, we have shown that the argument schemata that appear to ground typical conditional fallacy proofs mounted against alethic antirealism are valid in modal conditional logic only if certain semantic principles are accepted, which in turn rest on non-trivial metaphysical assumptions. The latter are: quasi transitivity of the accessibility relation given the similarity relation, and special containment of the similarity relation inside the accessibility relation. An immediate consequence of this is that the conditional fallacy - at least when attributed to the antirealist - is not just a fallacy of *reasoning*, but one that depends on questionable philosophical assumptions. The antirealist does not

appear committed to these two metaphysical assumptions. So long as her commitment to these assumptions is not convincingly demonstrated, no *proof* that the antirealist commits a conditional fallacy depending on III or IV is possible within modal conditional logic.

Although our findings are important and encouraging for the antirealist, they do not put an end to the antirealist's struggle with the conditional fallacy challenge. For even though the validity of III and IV does depend on the assumptions exposed in Theorems 3 and 4, there might be *variants* of III and IV that could commit the antirealist to a neighbouring conditional fallacy independent from these assumptions. Despite this, it seems to us that our paper supplies the antirealist with a quite general recipe for defusing conditional fallacy arguments: when threatened by one of these arguments, the antirealist should try to uncover its metaphysical assumptions and deny her commitment to them. This paper exemplifies two successful applications of this general strategy.

The focus of this paper has mainly been on alethic antirealism, but our semantic analysis of the conditional fallacy is utterly general and potentially bears on any use of the argument schemata III and IV. The falsities or absurdities deemed to follow from the analysis of *any* dispositional or response-dependence notion through versions of the conditional fallacy resting on any of these argument schemata will be defused whenever one of the metaphysical principles grounding their validity cannot be assumed.

## References

- [1] David Malet Armstrong. *A Combinatorial Theory of Possibility*. Cambridge, 1989.
- [2] Daniel Bonevac, Josh Denver, and David Sosa. The conditional fallacy. *Philosophical Review*, 115:273–316, 2006.
- [3] Berit Brogaard and Joe Salerno. Anti-realism, theism and the conditional fallacy. *Nous*, 39:123–129, 2005.
- [4] John P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22(1):76–84, January 1981.
- [5] John Divers. *Possible Worlds*. Routledge, London, 2002.
- [6] Michael Dummett. *The Seas of Language*. OUP, 1993.
- [7] Eline B. Gundersen. The chamaleon's revenge. response-dependence, finks and provisoed biconditionals. *Philosophical Studies*, 153:435–441, 2011.
- [8] Mark Johnston. How to speak of the colors. *Philosophical Studies*, 68:221–263, 1992.
- [9] David Lewis. *Counterfactuals*. Harvard University Press, 1973.

- [10] Edwin .D. Mares. Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic*, 38:516–526, 1997.
- [11] C. B. Martin. Dispositions and conditionals. *Philosophical Quartely*, 44:1–8, 1994.
- [12] Luca Moretti. Brogaard and Salerno on antirealism and the conditional fallacy. *Philosophical Studies*, 140(2):229–246, 2008.
- [13] Christopher Peacocke. Metaphysical necessity: Understanding, truth and epistemology. *Mind*, 106:521–574, 1997.
- [14] Alvin Plantinga. How to be an anti-realist. In *Proceedings and addresses of the american philosophical association*, volume 56, pages 47–70, 1982.
- [15] Philip Quinn. Metaphysical necessity and modal logic. *The Monist*, 65:444–455, 1982.
- [16] Michael Rea. Theism and epistemic truth-equivalences. *Nous*, 34:291–301, 2000.
- [17] Nathan Salmon. The logic of what might have been. *Philosophical Review*, 98:3–34, 1989.
- [18] Nathan Salmon. This side of paradox. *Philosophical Topics*, 21:187–197, 1993.
- [19] Robert K. Shope. The conditional fallacy in contemporary philosophy. *The Journal of Philosophy*, 75(8):397–413, August 1978.
- [20] Robert Stalnaker. A theory of conditionals. In Nicholas Resher, editor, *Studies in Logical Theory*, pages 98–112. OUP, 1968.
- [21] Frank Veltman. *Logic for conditionals*. PhD thesis, Department of Philosophy, University of Amsterdam, 1985.
- [22] Ralph Wedgwood. The price of non-reductive physicalism. *Nous*, 34:400–421, 2000.
- [23] Crispin Wright. *Truth and Objectivity*. Cambridge University Press, 1992.
- [24] Crispin Wright. Truth as sort of epistemic: Putnam's peregrinations. *Journal of Philosophy*, 97(6):335–364, 2000.