

Defusing the conditional fallacy: a semantical approach*

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

April 6, 2010

Abstract

The conditional fallacy is an argument deemed to entail absurd consequences for notions defined in terms of counterfactuals. For instance, the antirealist notion of truth is typically defined in the terms of what a rational thinker would believe if she were suitably informed. This notion is deemed to entail, through the conditional fallacy, the absurdity that there is necessarily a rational thinker. If this were the case, alethic antirealism should probably be rejected. In this paper we analyse the conditional fallacy from a semantical (i.e. model-theoretic) point of view. This allows us to identify with precision the philosophical commitments that ground the validity of this argument. We show that the conditional fallacy is generally valid only if some non-compulsory and questionable 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. Though we focus primarily on alethic antirealism, our analysis generalises to other philosophical uses of the conditional fallacy.

1 Introduction: the Conditional Fallacy and Alethic Antirealism

Generally speaking, the conditional fallacy is an argument (or a family of arguments) deemed to *entail* paradoxical or absurd consequences for notions characterized in terms of *dispositional* properties expressed by a counterfactual.¹

In this paper by the ‘conditional fallacy’ we will mean any instance of the

*We are very grateful to Fred Kroon, Jonathan McKeown-Green, Olivier Roy, Jeremy Seligman and Elia Zardini for valuable discussions and important criticisms upon drafts of this paper.

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

following argument form, which we call CF:

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

In the instances of CF, the first premise typically provides a definition of a notion in terms of a counterfactual property. The second premise states that a specific modal relation holds between the definiendum, p , and a *component* of the definiens $q \Box \rightarrow r$. The problematic consequences entailed by CF would depend on the fact that whenever the antecedent of the counterfactual, q , in the first premise necessarily implies the definiendum, the latter proves necessarily true. For example, suppose p says ‘ x is soluble’, q says ‘ x is immersed in water’, and r says ‘ x dissolves’. $\Box(p \equiv (q \Box \rightarrow r))$ provides a definition of the notion of solubility given in terms of the disposition to dissolve in water. This definition might appear straightforward. Suppose however that x is identical to the empty set. In this case, q is presumably *necessarily false*, which implies that $\Box(q \supset p)$. Hence, by CF, we derive $\Box p$ - i.e. necessarily, the empty set is soluble - which is false and absurd.² Hence, a metaphysician who wanted to make use of dispositional definitions like those instantiated by the first premise of CF should argue that CF is invalid. One of our purposes here is to show that whether or not CF is valid basically depends on *metaphysical* reasons rather than mere logic.

As can be seen from the above example, but also more generally by the structure of CF, the conditional fallacy arises as a consequence of the interplay between modalities and conditionals. In this paper, we investigate the relationship between modalities and conditionals and show that the conditional fallacy is not as pervasive and inescapable as has been argued before (for instance by Brogaard and Salerno [3]). We uncover crucial metaphysical assumptions that provide both sufficient and necessary conditions for the validity of CF.

The conditional fallacy has been argued to imply paradoxical consequences for *alethic antirealism*. In this paper we will primarily focus on these alleged consequences, though - we believe - our analysis could easily generalize to other philosophical uses of the conditional fallacy. The alethic antirealist analyzes the notion of truth in terms of epistemic notions such as justification or rational acceptability. The alethic antirealist typically endorses 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 ideal (or sufficiently good) to determine whether x ’, and $R(x)$ means ‘it is rationally believed that x ’.

AR says that, necessarily, it is true that x if and only if, if epistemic conditions were ideal (or sufficiently good) to determine whether x , it would rationally be

²Similar examples can be found for instance in Fara[7], Bird[2], Martin[11], Johnston [8] and [9], Shope[19].

believed that x . Brogaard and Salerno [3] contend that antirealism resting on AR is flawed because it commits a conditional fallacy by entailing the absurdity that there is *necessarily* an epistemic agent. Consider the case in which x is the proposition $p =$ ‘epistemic conditions are ideal for determining whether some statement is true’. Given that $\Box(T(p) \equiv p)$, the alethic antirealist who accepts AR is committed to $\Box(p \equiv (Q(p) \Box \rightarrow R(p)))$. But the antirealist is also committed to $\Box(Q(p) \supset p)$ because, in this interpretation of p and $Q(p)$, $Q(p)$ entails p .³ $\Box(p \equiv (Q(p) \Box \rightarrow R(p)))$ and $\Box(Q(p) \supset p)$ are the two premises of an instance of CF that concludes with $\Box p$. So the antirealist would be committed to $\Box p$ - namely, to the claim that, necessarily, epistemic conditions are ideal for determining whether some statement is true. These epistemic conditions ‘include the existence of a properly placed epistemic agent. So, necessarily, there is an epistemic agent’ ([3], p.135).⁴

Brogaard and Salerno’s proof of CF is not completely original but improves upon earlier demonstrations given by Plantinga [14], Wright [23] and Rea [16].⁵ Plantinga [14] and Rea [16] take the conclusions of their proofs to show that the alethic antirealist is committed to some form of *theism* (i.e. to the thesis that, roughly, there is necessarily an omniscient epistemic agent/community). Wright [23] and Brogaard and Salerno [3] claim that this is a misguided interpretation of those results, which should instead be seen as instances of the general problem of the conditional fallacy that plagues counterfactual analyses. We find this claim straightforward and we will not question it in our paper.

A feature common to the proofs cited above is that they all aim to *derive* (i.e. syntactically) the conclusion rather than showing that the argument form of CF is *valid*. In other words, CF is typically conceived of as a logical *inference* the alethic antirealist is committed to. This may encourage the erroneous belief that the 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 CF are exposed, CF appears questionable. For instance, Moretti [12] has pointed out that all these proofs hinge on the controversial assumption that a counterfactual is deducible from a correlated strict conditional, which can be rejected by the antirealist. Moretti has suggested that elementary reformulations of Brogaard and Salerno’s proof that do not rest on this assumption are incorrect or unsound or can be dismissed by an antirealist who has metaphysical reasons to reject a modal system as strong as S5 (or as its intuitionistic version). Moretti’s conclusion is that no proof that antirealism resting on a counterfactual analyses of truth commits a conditional fallacy has yet been delivered.

In this paper, we focus on none of the proposed proofs of CF but on CF itself.

³Since $\Box(T(p) \equiv p)$ - i.e. ‘epistemic conditions are ideal for determining whether p ’ entails ‘epistemic conditions are ideal for determining whether *it is true* that p ’. From the latter, it follows that p - i.e. ‘epistemic conditions are ideal for determining whether some proposition is true’.

⁴We do not question the claim that the existence of ideal epistemic conditions implies the existence of an ideal agent, though this might be questioned.

⁵These demonstrations are found in various ways defective by Brogaard and Salerno. We believe that Brogaard and Salerno’s objections are basically correct.

We explicitly take a *semantical* approach; this will enable us to identify with great accuracy the semantical and thus metaphysical commitments that ground the validity of CF. We will show that CF is valid *only if* specific and apparently non-compulsory metaphysical assumptions are made. We will suggest that the alethic antirealist who accepts AR is not committed to at least two of 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 will be using in the paper. In Section 3 we introduce the semantical (or metaphysical) principles presupposed by the validity of CF. In Section 4 we show that the joint truth of these principles suffices to make CF valid. In Section 5 we apply modal correspondence theory to single out the formulae that correspond to these semantical principles. In Section 6, we show that dropping any of the semantical principles invalidates CF. In Section 7 we suggest metaphysical reasons to doubt the soundness of the semantical principles presupposed by CF. In Section 8 we draw the conclusion of the paper. Some proofs are left to the appendix.

2 Modal Conditional Logic

Let us call *Modal Conditional Logic*, denoted MCL, any modal logic enriched with logical resources for conditionals. The language of MCL, which we will refer to as \mathcal{L}_{MCL} , is based on a set PROP of propositional variables which can be combined recursively using the propositional connectives \sim and \vee , and the modalities \Box and $\Box\rightarrow$. We call the modality \Box the *necessity operator*, and the modality $\Box\rightarrow$ the *conditional operator*.⁶ The remaining connectives are defined in the usual way. For instance, the material conditional $\varphi \supset \psi$ is defined as $\sim\varphi \vee \psi$, the *possibility operator* $\Diamond\varphi$ by $\sim\Box\sim\varphi$, and the *might conditional* $\varphi \Diamond\rightarrow \psi$ by $\sim(\varphi \Box\rightarrow \sim\psi)$.

The most common semantics for conditional logic is the one introduced by Lewis in [10] and is represented by systems of spheres, as in Figure 1. Formally, a system of spheres is a partial order \leq_w defined over the set of worlds.⁷ We write $u \leq_w v$ and say that “ u is at least as similar to w as v is”. Lewis imposes a further assumption on the system of spheres, the so-called *limit assumption*, which states that the strict sub-relation $<_w$ of \leq_w is well-founded.⁸ In a system

⁶We choose the term ‘conditional’ as a neutral term standing for counterfactual operators. It is typical to refer to $\Box\rightarrow$ as a *conditional operator* in a logical context and as a *counterfactual operator* in a metaphysical context.

⁷A relation is a partial order if it is:

1. Reflexive: $\forall x : xRx$,
2. Antisymmetric: $\forall x, y : (xRy \ \& \ yRx) \Rightarrow x = y$, and
3. Transitive: $\forall x, y, z : (xRy \ \& \ yRz) \Rightarrow xRz$.

⁸The strict subrelation $<_w$ is define by: $u <_w v$ iff $u \leq_w v$ & $v \not\leq_w u$. A relation R is well-founded on a set W if for every non-empty subset W' of W , there exists an element $w \in W'$ such that not vRw for every $v \in W'$. Equivalently, a relation is well-founded if it does not contain infinite descending chains.

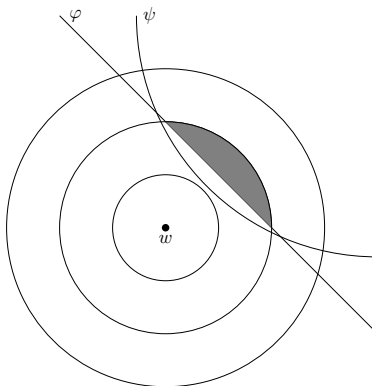


Figure 1: Graphical representation of the similarity operator with the limit assumption as a system of spheres.

of sphere, the limit assumption guarantees that for any world w , there is always (at least) one world which is most similar to w . The worlds that are most similar to w in a system of spheres are called the *minimal worlds* or the *closest worlds*. According to Lewis's semantics for conditional logic, a conditional is true in a world u if and only if its consequent is true in all possible worlds most similar to u in which its antecedent is true. Schematically:

$$\mathfrak{M}, u \models \varphi \Box \rightarrow \psi \quad \text{iff} \quad \mathfrak{M}, v \models \psi \text{ for every world } v \text{ minimal in } \leq_u \text{ such that } \mathfrak{M}, v \models \varphi.$$

In Figure 1, the shaded region represents the minimal φ -worlds. Each of these minimal φ -world is also a ψ -world, so $\mathfrak{M}, w \models \varphi \Box \rightarrow \psi$.

The limit assumption has been the source of much disagreement. A common example of a counterfactual situation not satisfying it, first given by Lewis himself in [10], is a world in which you are slightly taller than you actually are. There is no world most similar to the actual world in which your are slightly taller than you actually are, since for every world u in which you are taller than you actually are, you can always find a world in which you are still slightly taller than your actually are, yet shorter than in u . This implies that counterfactuals such as "If you were slightly taller, you would be a cinema" become vacuously true, contrary to expectations.

We take this criticism seriously, but also for the sake of generality, we have chosen a very general semantical definition for the similarity operator $\Box \rightarrow$. This semantics was first investigated in Veltman [21] and Burgess [4] and was motivated by an analysis of conditionals in which $\varphi \Box \rightarrow \psi$ is true if the $\varphi \& \sim \psi$ -worlds are more similar to the actual world than the $\sim \varphi \& \psi$ -worlds. We take this semantical definition to be minimal in the same way that an accessibility relation R with no assumption is minimal with respect to \Box - and yields the logic known as K . In the remainder of the paper, we work with the minimal modal conditional logic, since we want to isolate the crucial assumptions yielding the

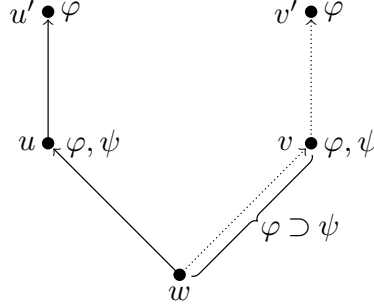


Figure 2: Representation of the *MCL* semantics for $\varphi \Box \rightarrow \psi$. Dotted arrows stand for (possibly) infinite paths between two worlds.

conditional fallacy. The limit assumption is not one of them, and we support this claim by working in systems in which it need not be assumed.⁹

Models for minimal *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 the relation R the *accessibility relation* and \leq the *similarity relation*.

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 - $V(p)$ is the set of worlds in which p is true. We base our interpretation of arbitrary formulae of the language on this propositional valuation and we define the notion of *satisfaction at a world in a model*, written $\mathfrak{M}, w \models \varphi$, recursively:

$$\begin{aligned}
\mathfrak{M}, w \models p & \quad \text{iff} \quad w \in V(p) \\
\mathfrak{M}, w \models \sim \varphi & \quad \text{iff} \quad \mathfrak{M}, w \not\models \varphi \\
\mathfrak{M}, w \models \varphi \vee \psi & \quad \text{iff} \quad \mathfrak{M}, w \models \varphi \text{ or } \mathfrak{M}, w \models \psi \\
\mathfrak{M}, w \models \Box \varphi & \quad \text{iff} \quad \forall u \text{ s.t. } wRu, \mathfrak{M}, u \models \varphi \\
\mathfrak{M}, w \models \varphi \Box \rightarrow \psi & \quad \text{iff} \quad \forall u \in W \text{ s.t. } \mathfrak{M}, u \models \varphi, \exists u' \text{ s.t.} \\
& \quad 1) u' \leq_w u, \\
& \quad 2) \mathfrak{M}, u' \models \varphi \text{ and} \\
& \quad 3) \forall u'' : u'' \leq_w u' \Rightarrow \mathfrak{M}, u'' \models \varphi \supset \psi
\end{aligned}$$

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

⁹Should a metaphysician also choose to work in more economical systems!

The reader may read the semantics for $\varphi \Box \rightarrow \psi$ as “a conditional is true in a world w iff in all worlds *sufficiently* similar to w in which its antecedent is true, its consequent is also true”.

The semantics of $\varphi \Box \rightarrow \psi$ is illustrated in Figure 2. In that picture, a black arrow between two worlds s and t indicates that $s \leq_w t$, so for instance, $u \leq_w u'$. Dotted arrows stand for possibly infinite paths between two worlds. For instance, the dotted arrow between v and v' indicates that there are worlds $s_1, s_2, \dots, s_n, \dots$ (possibly infinitely many) such that $v \leq_w s_1, s_1 \leq_w s_2, s_2 \leq_w s_3, \dots, s_i \leq_w s_{i+1}, \dots$ and for every $i, s_i \leq_w v'$. For readability, we have not drawn reflexive and transitive arrows, although they are implicitly there, because of assumptions on the similarity models. In this model, it is the case that $\mathfrak{M}, w \models \varphi \Box \rightarrow \psi$, since for every world $s \in W_w$ such that $\mathfrak{M}, s \models \varphi$, there is a world t such that $\mathfrak{M}, t \models \varphi$, and for every world t' with $t' \leq_w t$, $\mathfrak{M}, t' \models \varphi \supset \psi$. For instance, $\mathfrak{M}, u' \models \varphi$, but $u \leq_w u'$ and $\mathfrak{M}, u \models \varphi$, and every world s such that $s \leq_w u$ (namely u and w) is such that $\mathfrak{M}, s \models \varphi \supset \psi$. Likewise for world v' , v is such that $v \leq_w v'$, $\mathfrak{M}, v' \models \varphi$ and for every world s such that $s \leq_w v$ (here possibly infinitely many, including w), $\mathfrak{M}, s \models \varphi \supset \psi$.

If one adds the limit assumption to the semantics for $\Box \rightarrow$ then one recovers the Lewis semantics.¹⁰ We will be working with minimal conditional logic for the for the rest of the paper. Let us now investigate the assumptions needed to validate the conditional fallacy.

3 Three Crucial Assumptions

As we show below, CF is valid if and only if three crucial assumptions about *MCL* hold true; precisely: 1) transitivity of R , 2) weak centering of \leq_w and 3) relation containment of \leq_w inside R .

Transitivity of R is well-known. It simply states that if a world v is related to a world u via R , and a world w is related to a world v via R , then also w is related to u through R . Formally: $(Rwv \ \& \ Rvu) \Rightarrow Rwu$.

Weak centering is the assumption that no world can be closer or more similar to a world w than w itself. Formally: $w \leq_w w$ for every world w and there is no world $v \neq w$ such that $v <_w w$. 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 an agent might be misguided about the real world and think she lives in a world metaphysically far remote from the real world. 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. (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, $w \leq_w w$ for every world w and for every world v different from w , $w <_w v$).

Relation containment of \leq_w inside R is the assumption that any two worlds which are related by \leq_w are also accessible from each other via R . The formal

¹⁰See Appendix, Proposition 5.

definition is more illuminating: For every world w, u and v , if $u \leq_w v$, then both Rwu and Rwv . This means that any world that can be compared in the similarity relation is an accessible world. We do not know of a common name to refer to this assumption of \leq being contained in R . A slogan that might help the reader remember the content of the assumption is that “conditionality entails possibility”.

Weak centering and transitivity are assumptions that can be made and motivated in isolation, but when combined with relation containment in a multimodal context, further assumptions are forced as consequences of the interaction between the relations. Straightforward consequences are that if an accessibility relation R_1 is contained in an accessibility relation R_2 , and R_1 is either reflexive or total, then R_2 must also be reflexive¹¹ or total,¹² respectively. In the context of the present paper, since \leq_w is assumed to be a reflexive relation, then so must be R . A less obvious consequence of relation containment is that it prevents \leq_w from being connected (for every pair of worlds u and v , either $u = v$, $u \leq_w v$ or $v \leq_w u$) if R is not transitive.¹³ That reflexivity of R follows as a consequence of relation containment is not perplexing in our opinion, especially in light of correspondence theory discussed in Section 5.¹⁴ The fact that relation containment prevents \leq_w to be connected if R is not transitive is more important and forces the metaphysician to assume transitivity of R if she is already committed to relation containment and to connectedness of \leq_w . But as we will suggest in Section 7, there may be metaphysical reasons to reject transitivity, in which case one also needs to drop connectedness of \leq_w in order to keep relation containment - or drop relation containment in order to keep connectedness.¹⁵

4 The Proof

Let us now show that weak centering, relation containment and transitivity of R constitute jointly a sufficient condition for the conditional fallacy.

Theorem 1 *In a conditional modal logic with 1) weak centering, 2) relation containment and 3) transitivity of R , the conditional fallacy is valid.*

Proof: We show that CF is valid:

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

¹¹See Appendix, Proposition 6

¹²See Appendix, Proposition 7

¹³See Appendix, Proposition 8.

¹⁴Note also that reflexivity is axiomatized by $\Box\varphi \supset \varphi$, i.e., if φ is necessary, then φ is true, which is expected in a metaphysical context.

¹⁵When combined with the limit assumption, connectedness implies uniqueness of closest worlds. For a rejection of this latter assumption, see Lewis [10], p.79-83, on *conditional excluded middle*.

Assume that a) $\mathfrak{M}, w \models \Box(p \equiv (q \Box \rightarrow r))$, b) $\mathfrak{M}, w \models \Box(q \supset p)$, but that c) $\mathfrak{M}, w \not\models \Box p$. By the latter assumption, $\mathfrak{M}, w \models \sim \Box p$, which is equivalent by definition to $\mathfrak{M}, w \models \Diamond \sim p$. By the semantical definition of $\Box \rightarrow$, this implies that there is a v such that Rwv and $\mathfrak{M}, v \models \sim p$. By assumption a), also $\mathfrak{M}, v \models p \equiv (q \Box \rightarrow r)$, so $\mathfrak{M}, v \not\models q \Box \rightarrow r$. By the semantical definition, this implies that there exists a $v' \in W$ such that $\mathfrak{M}, v' \models q$. Now, for every u , if $u \leq_v v'$ and $\mathfrak{M}, u \models q$, then there exists a u' such that $u' \leq_v u$ and $\mathfrak{M}, u' \not\models q \supset r$, so $\mathfrak{M}, u' \models q$ and $\mathfrak{M}, u' \not\models r$. For any such u' , $u' \leq_v u$ implies that Rvu' by **relation containment**. Since Rwv , also Rwu' , by **transitivity of R**. By assumption b), $\mathfrak{M}, u' \models q \supset p$, so $\mathfrak{M}, u' \models p$, by Modus Ponens. Finally, by assumption a), $\mathfrak{M}, u' \models p \equiv (q \Box \rightarrow r)$, so $\mathfrak{M}, u' \models q \Box \rightarrow r$. Hence, by **weak centering**, $\mathfrak{M}, u' \models r$. This contradicts the claim that $\mathfrak{M}, u' \not\models r$, derived before. Therefore, it must be that $\mathfrak{M}, w \models \Box p$, as desired. ♣

5 Modal Correspondence Theory

Modal logic is a great device to talk about relational structures. In metaphysics of modality, the kind of relational structures that are investigated are sets of possible worlds which are related to one another by the accessibility relation. The latter is universal if any world is possible with respect to any other world; but sometimes it is more restricted, when metaphysicians develop a more constrained theory of modality.

The semantics of modal logic is essentially local, in that formulae are evaluated at worlds in models. A formula such as $\Box p$ may thus be true at certain worlds but false at others. One can however evaluate a modal formula from a more general perspective and ask whether it is *valid* in a specific *class* of models (true at every world in every model of the specific class) or in a specific *class* of frames (true at every world in every model based on any frame of that specific class). This distinction in perspective is one of the great advantages of modal logic. For modal logic can be used to express, on the one hand, local or first-order properties of models and, on the other hand, global or second-order properties. The branch of modal logic called *modal correspondence theory* investigates these second-order properties (cf., [20]). More accurately, it investigates which formulae or axioms correspond to which properties of models or frames. For example, it is well-known that $\Box \varphi \supset \varphi$ corresponds to the reflexivity of R and that $\Box \varphi \supset \Box \Box \varphi$ corresponds to its transitivity.

For the purpose of our paper it is important to answer the following question: are there formulae of MCL that correspond to the properties of weak centering and relation containment? That is, for any class of frames F , are there formulae φ and ψ such that $F \models \varphi$ if and only if every frame in F is weakly centered and $F \models \psi$ if and only if relation containment holds for every frame in F ?

The answer is “Yes!”. Here they are:

$$\begin{aligned} (p \& (p \Box \rightarrow q)) \supset q & \quad (\text{WC}) \\ \Box(p \supset q) \supset (p \Box \rightarrow q) & \quad (\text{RC}) \end{aligned}$$

It is well known in the literature on conditional logic that WC corresponds to weak centering (cf., for instance [10]). For a proof that RC corresponds to relation containment, see the Appendix, Theorem 9.

For the sake of generality, we have proved Theorem 9 without the limit assumption, following the minimal semantical definition of $\Box \rightarrow$. If one assumes the limit assumption, thus simplifying the semantics of the similarity operator to the more usual Lewis conditional represented as a system of sphere (see Figure 1), then Theorem 9 correspondingly simplifies to the following:

For any class of frames F , WC is valid in F if and only if any world v of F contained in a sphere centered around a world u is accessible from u .

It is typical in the literature on CF to work proof-theoretically with WC and RC i.e. the tendency is to take them to be axioms of inference, or to replace WC and RC with corresponding rules of inference (see for instance [3]). This muddies the philosophical waters. What modal correspondence theory shows is that these principles correspond to precise model theoretical assumptions. When one applies logic in a metaphysical context, it becomes apparent that these model theoretical assumptions are in fact metaphysical assumptions whose soundness and acceptability (if any) is largely independent from logic. Hence, if one wants to appeal to the validity of instances of CF to conclude that, say, antirealism is paradoxical, one has to offer, first of all, a *metaphysical* defense of weak centering and relation containment.

6 Defusing the Conditional Fallacy

Model-theory can be used to understand to what extent weak centering, relation containment and transitivity of R are crucial in validating CF. Indeed, we will now show that dropping any of these three assumptions *invalidates* CF. Our strategy is the standard model-theoretical one used to prove that an argument form is invalid: that is, we provide a counterexample. We do it in three possible configurations, in each of which one of the assumptions is dropped but the others are retained. In each case CF fails. Here our point is not metaphysical but just logical: if 4 or WC or RC are not assumed to be valid, then CF is invalid .

Theorem 2 *If transitivity of R is not assumed, then CF is invalid.*

Proof: Figure 3 provides a counterexample. In that picture, a black arrow is drawn between two worlds u and v if uRv . Since uRv and vRw but not uRw , we have a failure of transitivity. In the model, each world x is such that $\mathfrak{M}, x \models q \supset p$. Hence, $\mathfrak{M}, u \models \Box(q \supset p)$. Now, $\mathfrak{M}, u \models \sim(q \Box \rightarrow r)$, since w is the minimal q -world in \leq_u and $\mathfrak{M}, w \models \sim r$. Since $\mathfrak{M}, u \models \sim p$, we have that $\mathfrak{M}, u \models (p \equiv (q \Box \rightarrow r))$. For the same reason, $\mathfrak{M}, v \models (p \equiv (q \Box \rightarrow r))$. Since u and v are the only worlds x such that uRx , $\mathfrak{M}, u \models \Box(p \equiv (q \Box \rightarrow r))$. Finally, since $\mathfrak{M}, v \models \sim p$ and uRv , $\mathfrak{M}, u \not\models \Box p$. Therefore, CF is invalid. ♣

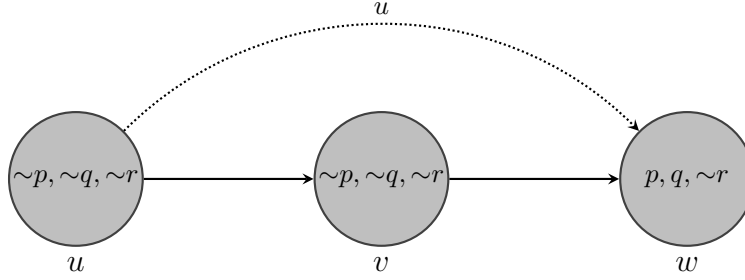


Figure 3: A counterexample to CF when transitivity of R is not assumed. Black and dotted arrows stand for accessibility and similarity relations respectively. Reflexive arrows are left implicit.

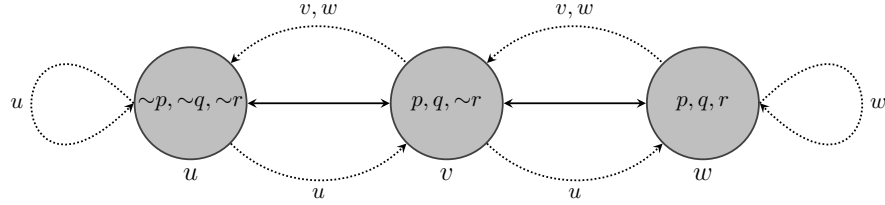


Figure 4: A counterexample to CF when weak centering is not assumed. Black and dotted arrows stand for accessibility and similarity relations respectively. In this figure, the reflexive links for R are not drawn, but implicitly assumed, whereas reflexive links for \leq only hold if they are drawn.

Theorem 3 *If weak centering is not assumed, then CF is invalid.*

Proof: Figure 4 provides a counterexample. In the model, reflexive and transitive accessibility links are implicitly assumed, but not drawn. Reflexive similarity links, however, only obtain when drawn. In the model, every world x is such that $\mathfrak{M}, x \models q \supset p$, so $\mathfrak{M}, u \models \Box(q \supset p)$. Since v is the minimal q -world in \leq_u and $\mathfrak{M}, v \not\models r$, $\mathfrak{M}, u \not\models (q \Box \rightarrow r)$. But $\mathfrak{M}, u \not\models p$, so $\mathfrak{M}, u \models (p \equiv (q \Box \rightarrow r))$. And since w is the minimal q -world in \leq_v and $\mathfrak{M}, w \models r$, $\mathfrak{M}, v \models (q \Box \rightarrow r)$, so $\mathfrak{M}, v \models (p \equiv (q \Box \rightarrow r))$. Finally, w is the minimal q -world in \leq_w and $\mathfrak{M}, w \models r$, so $\mathfrak{M}, w \models (q \Box \rightarrow r)$. Since $\mathfrak{M}, w \models p$, we get that $\mathfrak{M}, w \models (p \equiv (q \Box \rightarrow r))$. Hence, each world x in the model is such that $\mathfrak{M}, x \models (p \equiv (q \Box \rightarrow r))$. Therefore, $\mathfrak{M}, u \models \Box(p \equiv (q \Box \rightarrow r))$. Finally, $\mathfrak{M}, v \not\models p$ and uRv , so $\mathfrak{M}, u \not\models \Box p$. Therefore, CF is invalid. ♣

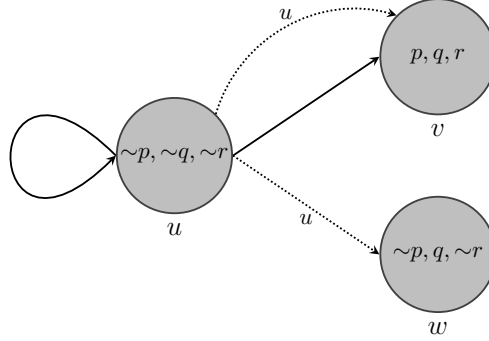


Figure 5: A simple counterexample to CF when relation containment is not assumed. Black and dotted arrows stand for accessibility and similarity relations respectively. In this figure, the reflexive links for R only hold if they are drawn, whereas reflexive links for \leq hold implicitly, but are not drawn.

Theorem 4 *If relation containment is not assumed, then CF is invalid.*

Proof: Figure 5 provides a counterexample. In this case, reflexivity of R need not be assumed.¹⁶ Accordingly, we draw reflexive R arrow explicitly if they hold. Each world x with uRx (namely u and v) is such that $\mathfrak{M}, x \models q \supset p$, so $\mathfrak{M}, u \models \Box(q \supset p)$. Since w is a minimal q -world in \leq_u and $\mathfrak{M}, w \not\models r$, $\mathfrak{M}, u \not\models (q \Box \rightarrow r)$. But $\mathfrak{M}, u \not\models p$, so $\mathfrak{M}, u \models (p \equiv (q \Box \rightarrow r))$. Also v is a minimal q -world according to \leq_v and $\mathfrak{M}, v \models r$, so $\mathfrak{M}, v \models (q \Box \rightarrow r)$. Since $\mathfrak{M}, v \models p$, we have that $\mathfrak{M}, v \models (p \equiv (q \Box \rightarrow r))$. Hence, $\mathfrak{M}, u \models \Box(p \equiv (q \Box \rightarrow r))$. But $\mathfrak{M}, u \not\models \Box p$, since $\mathfrak{M}, u \models \sim p$ and uRu . Therefore, CF is invalid. ♣

A variant of the conditional fallacy investigated in Moretti [12], which we call the *weak conditional fallacy*, is given in the following argument form:

$$\begin{array}{l}
 \Box(p \equiv (q \Box \rightarrow r)) \\
 \Box(q \supset p) \\
 \Diamond q \\
 \hline
 \therefore \Box p
 \end{array}$$

This version adds that premise $\Diamond q$ to discard cases of the fallacy in which q is not possible. There is no consensus on how to treat conditionals with impossible antecedents, and this version of the fallacy avoids the debate with the introduction of $\Diamond q$ as a premise. Moretti [12] suggests that the weak conditional fallacy is the one implicitly endorsed by Brogaard and Salerno in [3]. Even though we have not made this point explicitly, we have constructed our counter-examples in Theorems 2-4 so that they can be applied against the weak conditional fallacy too, as the reader can check.

¹⁶Since the conditions of Proposition 6 (see the Appendix) are not satisfied.

What philosophical significance do these theorems have? Suppose the antirealist sticks to classical logic. The antirealist who accepts AR would commit a conditional fallacy, and would thus run afoul a reductio ad absurdum, *only if* she also accepted transitivity of R , weak centering of \leq and relation containment. Consider for instance relation containment. Why should the antirealist qua antirealist maintain that the similarity relation \leq is a subset of the accessibility relation R ? Whoever argues in favor of a conditional fallacy committed by 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.

The reader might be concerned that 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 2-4 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 conditionals and explore how our counter-examples could be adapted to such system, but we leave this not easy task to the reader, as this would distract us from our main topic.

It is also worth noticing that though the focus of this paper is mainly on alethic antirealism, the semantical analysis of the conditional fallacy put forward here is utterly general and potentially bears on any use of this argument. The absurdities deemed to follow from the analysis of *any* dispositional notion via the conditional fallacy will be defused whenever one of the three metaphysical principles grounding the validity of the conditional fallacy cannot be assumed.

7 Metaphysical reasons to reject transitivity of R and relation containment

In contemporary philosophy there is no agreement on the correct analysis of metaphysical modality - especially on the properties of the accessibility relation R . The interesting debate on this topic is still open and on going. Before concluding we would like to suggest two arguments that the antirealist could use to reject transitivity of R or relation containment. Nathan Salmon[17, 18] has raised a Sorites argument against the transitivity of R . Armstrong [1], while defending a *combinatorialist* theory of possibility, has found a reason to reject the symmetry of R . If one accepts combinatorialism -at least the variant of it presented in Armstrong [1] one can conceive of frames including possible worlds that are non-mutually accessible. Since some non-mutually accessible worlds can be *closest* to one another, \leq turns out *not* to be included in R . Thus, if combinatorialism is true, relation containment should be rejected. We suggest that the antirealist might exploit versions of either argument to reject the conditional fallacy.

This is Salmon's argument in more detail: consider an artifact for example a table, which we will call T. It is intuitive that T, while retaining its numerical identity, could have originated from a piece of tree trunk W1 very slightly different from the piece of tree trunk W0 from which T has actually originated. Suppose for instance that W1 has the same shape and size as W0 but is taken from one millimeter further down the same trunk as W0. In short, it is *possible* that T is made of W1. Consider now that if T had actually originated from W1, it would plausibly be true that T could have originated from W2 - a piece of wood taken from one additional millimeter further down the same trunk. Thus it is possible that it is possible that T is made of W2. 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 it is possible ... that T is made of W1000, namely, a piece of wood that differs from T's actual piece of wood by one meter. If R is transitive, we should derive that it is just possible that T is made of W1000. But this seems incorrect: if T had originated from a piece of wood that differs from T's actual piece of wood by one meter, T would plausibly be a distinct individual!

We are not endorsing Salmon's argument. Our point is simply that there is apparently no reason why the alethic antirealist could not endorse Salmon's argument when threatened by the conditional fallacy objection.

This is Armstrong's argument in more detail: a combinatorialist theory of possibility is a metaphysical picture according to which possible worlds are rearrangements or recombinations of contingent elements, e.g. individuals and properties existing in the actual world. For the combinatorialist, a state of affairs is possible if and only if it can be obtained recombining actually existing simple individuals and actually instantiated simple properties. 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. (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 - 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. This constitutes a genuine possibility for the combinatorialist - something that could have happened. Possible worlds in which a does not exist are accessible from the actual world. Consider one of these worlds - let us call it w . The actual world is not accessible from w because it cannot be obtained by re-arranging the contingent elements constituting w , *which do not include a* . Hence, for the combinatorialist, the accessibility relation R is not symmetric. Suppose now that w is the same as the actual world, except for the fact that w does not include a . This is *prima facie* conceivable. Clearly, the actual world and w are *extremely* similar, but the former is not accessible from the latter. Hence, for the combinatorialist, \leq is not included in R .

Combinatorialism is affected by difficulties, but this is true of any other

interesting theory of possibility. We are not defending combinatorialism. Our point is that the alethic 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*. Consequently, the alethic antirealist can reject relation containment, and thus CF.

There are other interesting arguments against the symmetry of R which cannot be reviewed here - see for instance, among many others, Wedgwood[22], Peacocke[13], Dummett ([6]: 328-348) and Quinn[15].¹⁷ Some of these arguments are less general than Armstrong's and Salmon's, as they depend on embracing specific views in particular areas of philosophy. For example, Wedgwood ([22]) has made a powerful case that nonreductive physicalism about mental properties - a view quite fashionable today - cannot be true if R is symmetric. It seems to us that even "local" arguments of this type can shed doubts on the symmetry of R . Versions of these arguments might perhaps be adduced by the antirealist when pressed by the conditional fallacy objection.

8 Conclusion

We have shown that CF proves valid only if certain semantical principles are accepted, which in turn rest on genuine metaphysical assumptions; namely: transitivity of the accessibility relation, weak centering of the similarity relation, and relation containment of the similarity relation inside the accessibility relation. An immediate consequence of this is that the conditional fallacy is not just a fallacy of *reasoning*, but one that depends on substantive and controversial philosophical assumptions. The moral we should draw is this: the metaphysician who intends to appeal to dispositional properties, or to define notions in terms of counterfactuals, may still do so in many cases. But she must make sure that her surrounding metaphysical commitments do not entail the three semantical principles necessary to CF. This is as much as logic alone can show in the philosophical debate about the conditional fallacy.

Let us come to the specific issue of alethic antirealism. We have suggested that the antirealist who accepts a counterfactual characterization of truth is not committed to at least two of the semantical principles necessary to CF. Moretti [12] has argued that no proof that the alethic antirealist commits a conditional fallacy has yet been delivered. We can now make that diagnosis more accurate: so long as the antirealist's commitment to the transitivity of R or the relation containment of R inside \leq stand undemonstrated, no proof of this sort is possible at all.

Appendix

Proposition 5 *The minimal semantics for $\Box \rightarrow$ augmented with the limit assumption is equivalent to the Lewis semantics.*

¹⁷We are grateful to Antti Keskinen and Ralf Wedgwood for drawing our attention to some of these papers.

Proof: Let us show first that the minimal semantics for $\Box \rightarrow$ augmented with the limit assumption entails the Lewis semantics. Assume that $\mathfrak{M}, w \models \varphi \Box \rightarrow \psi$, that \leq_w is well-founded and that there is a world $u \in W$ such that $\mathfrak{M}, u \models \varphi$. By the semantics definition, there is a world $u' \in W$ such that $u' \leq_w u$, $\mathfrak{M}, u' \models \varphi$ and for every $u'' \leq_w u'$, $\mathfrak{M}, u'' \models \varphi \supset \psi$. Consider the set $\Sigma = \{v : v \leq_w u', \mathfrak{M}, v \models \varphi\}$. Since \leq_w is well-founded, there is a minimal element in Σ . Let v_μ be minimal in Σ , then $\mathfrak{M}, v_\mu \models \varphi$ by definition of Σ . But $\mathfrak{M}, v_\mu \models \varphi \supset \psi$ by assumption, so $\mathfrak{M}, v_\mu \models \psi$. Hence, all minimal elements in Σ satisfy ψ . Since u was chosen arbitrarily, all minimal φ -worlds are ψ -worlds, the Lewis semantics for $\Box \rightarrow$.

Let us show that the Lewis semantics entails the minimal semantics for $\Box \rightarrow$ augmented with the limit assumption. Assume the Lewis semantics for $\varphi \Box \rightarrow \psi$ and assume that $\mathfrak{M}, w \models \varphi$. Let u be any world with $\mathfrak{M}, u \models \varphi$. By the limit assumption, there is a minimal world u' such that $u' \leq_w u$ and $\mathfrak{M}, u' \models \varphi$, so conditions 1) and 2) of the minimal semantics are satisfied. For the third condition, let u'' be any world such that $u'' \leq_w u'$. Since u' is a minimal world in \leq_w , u'' is also minimal in \leq_w (otherwise u' is not minimal, a contradiction). Hence, $\mathfrak{M}, u'' \models \varphi$. From the Lewis semantics, $\mathfrak{M}, u'' \models \psi$, so $\mathfrak{M}, u'' \models \varphi \supset \psi$. Therefore, $\forall u'' : u'' \leq_w u' \Rightarrow \mathfrak{M}, u'' \models \varphi \supset \psi$. ♣

Proposition 6 *Weak centering and relation containment entail Reflexivity of R .*

Proof: The proof is straightforward. Weak centering entails that $w \leq_w w$ for every w . Hence relation containment entails that wRw for every w . ♣

Proposition 7 *Weak centering and totality of \leq_w entail totality of R .*

Proof: The proof is also straightforward. Since for every two worlds u, v , either $u \leq_w v$ or $v \leq_w u$ weak centering entails that either uRv or vRu . ♣

Proposition 8 *Let R be some accessibility relation and let $\{\leq_w\}_{w \in W}$ be a collection of similarity relations (partial orders). Assume that for every world $w, u, v \in W$, if $u \leq_w v$, then both wRu and wRv (relation containment). If R is not assumed to be transitive, then it is not the case that \leq_w is connected (i.e. it is not the case that for every $u, v \in W, u = v, u \leq_w v$ or $v \leq_w u$) for every $w \in W$.*

Proof: Assume relation containment and assume that R is not transitive. Since R is not transitive, there is a model \mathfrak{M} and worlds u, v, w such that uRv, vRw but not uRw . Since \leq_u is reflexive, $u \leq_u u$ implies that uRu , so $u \neq w$. Suppose, that \leq_u is connected, then either $u \leq_u w$ or $w \leq_u u$, which imply that uRw by relation containment, a contradiction. Hence \leq_u cannot be a connected relation. ♣

Theorem 9 $\mathfrak{F} \models \Box(p \supset q) \supset (p \Box \rightarrow q)$ iff $\forall xyz \in W$: if $x \leq_z y$, then Rzx & Rzy .

Proof: From right to left, assume that $\forall xyz \in W(x \leq_z y \Rightarrow (Rzx \& Rzy))$ and assume that $\mathfrak{M}, u \models \Box(p \supset q)$ for some arbitrary model based on \mathfrak{F} . We need to show that $\mathfrak{M}, u \models (p \Box \rightarrow q)$. Take $v \in W$ arbitrary such that $\mathfrak{M}, v \models p$ and assume that there is no v' such that 1) $v' \leq_u v$, 2) $\mathfrak{M}, v' \models p$ and 3) $\forall v''(v'' \leq_u v' \Rightarrow \mathfrak{M}, v'' \models p \supset q)$. By simple logic manipulations, this is equivalent to $\forall v'(v' \leq_u v \& \mathfrak{M}, v' \models p \Rightarrow \exists v''(v'' \leq_u v' \& \mathfrak{M}, v'' \not\models p \supset q)$, call this (*). But since \leq_u is a reflexive relation, $v \leq_u v$ and $\mathfrak{M}, v \models p$ by assumption, so $\exists w(w \leq_u v \& \mathfrak{M}, w \not\models p \supset q)$ by (*). Finally, from our original assumption, $w \leq_u v$ implies that Ruw and Ruv , and since $\mathfrak{M}, u \models \Box(p \supset q)$, we get that $\mathfrak{M}, w \models p \supset q$, a contradiction. Therefore, for every v with $\mathfrak{M}, v \models p$, there exists a v' such that 1) $v' \leq_u v$, 2) $\mathfrak{M}, v' \models p$ and 3) $\forall v''(v'' \leq_u v' \Rightarrow \mathfrak{M}, v'' \models p \supset q)$, which implies, by the semantical definition of $\Box \rightarrow$, that $\mathfrak{M}, u \models (p \Box \rightarrow q)$.

From left to right, we argue by contraposition. Assume that $u \leq_w v$ but not $(Rwu \& Rvw)$. We need to show that $\mathfrak{F} \not\models \Box(p \supset q) \supset (p \Box \rightarrow q)$ and we show this by constructing a model \mathfrak{M} based on \mathfrak{F} such that $\mathfrak{M}, w \not\models \Box(p \supset q) \supset (p \Box \rightarrow q)$. We consider two cases: 1) $u \leq_w v$ and not Rwu , and 2) $u \leq_w v$ and not Rvw . In the first case, let $V(p) = \{u\}$ and $V(q) = \emptyset$. Then $\mathfrak{M}, w \models \Box(p \supset q)$, since for every world v' with Rwv' , $\mathfrak{M}, v' \not\models p$, which implies that $\mathfrak{M}, v' \models p \supset q$, but $\mathfrak{M}, w \not\models (p \Box \rightarrow q)$, since u is the only world such that $u \leq_w u$ and $\mathfrak{M}, u \models p$, but $\mathfrak{M}, u \not\models q$. In the second case, let $V(p) = \{v\}$ and $V(q) = \emptyset$. For similar reasons, $\mathfrak{M}, w \models \Box(p \supset q)$, but $\mathfrak{M}, w \not\models (p \Box \rightarrow q)$. Therefore, $\mathfrak{F} \models \Box(p \supset q) \supset (p \Box \rightarrow q)$. ♣

References

- [1] David Malet Armstrong. *A combinatorial theory of possibility*. Cambridge, 1989.
- [2] Alexander Bird. Dispositions and antidotes. *Philosophical Quarterly*, 48:227–234, 1998.
- [3] B. Brogaard and J. 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, Oxford, 1993.
- [7] Michael Fara. Dispositions and habituals. *Nous*, 43:43–82, 2005.
- [8] Mark Johnston. how to speak of the colors. *Philosophical Studies*, 68:221–263, 1992.
- [9] Mark Johnston. Objectivity refigured: Pragmatism and modern idealism. In Crispin Wright and John Haldane, editors, *Reality, Representation, and Projection*, page 85130. Oxford University Press, 1993.

- [10] David Lewis. *Counterfactuals*. Harvard University Press, 1973.
- [11] C. B. Martin. Dispositions and conditionals. *Philosophical Quarterly*, 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] Johan van Benthem. Correspondence theory. In D. Gabbay and F. Guenther, editors, *Handbook of philosophical logic, 2nd edition*, pages 325–408. Kluwer Academic Publishers, 2001.
- [21] Frank Veltman. *Logic for conditionals*. PhD thesis, Department of Philosophy, University of Amsterdam, 1985.
- [22] Ralf Wedgwood. The price of non-reductive physicalism. *Nous*, 34:400–421, 2000.
- [23] Crispin Wright. Truth as sort of epistemic: Putnam’s peregrinations. *Journal of philosophy*, 97(6):335–364, 2000.