PENULTIMATE DRAFT – PLEASE CITE THE PUBLISHED VERSION

In: *Model-Based Reasoning, Abductive Cognition, Creativity: Inferences and Models in Science, Logic, Language, and Technology*, Ippoliti E., Magnani L., Arfini S. (eds.), Springer, Cham, pp. 89–116 (2024).

Counterfactuals, Models, and Scientific Realism

Fabio Sterpetti

Department of Philosophy
Sapienza University of Rome
fabio.sterpetti@uniroma1.it

Abstract

Counterfactuals abound in science, especially when one deals with models. Some models, namely highly idealized models, have assumptions that are metaphysically impossible. This means that in science one has often to deal with counterpossibles. According to the standard semantics for counterfactuals, all counterpossibles are vacuously true. But scientific practice shows that counterpossibles are not always regarded as vacuously true by scientists. To do justice of the use of counterpossibles in science, some authors think that we should adopt a semantics that allows for impossible worlds. It seems difficult to reconcile the role played by highly idealized models in science with scientific realism. Nevertheless, some authors think that it is possible to provide a realist account of highly idealized models. In this view, scientific realism should not be interpreted as the claim that models aim to provide accurate representations of models' targets, rather it should be interpreted as the claim that models aim to provide true modal information about models' targets. This variant of scientific realism seems to imply a commitment to some form of modal realism. I develop an objection to that variant of scientific realism by elaborating on some arguments originally developed to show that there are unknowable facts.

Keywords: Model-based Reasoning; Counterfactual Reasoning; Scientific Realism; Modal Realism; Inconceivable Possibilities

1. Counterfactuals and model-based reasoning

There is an increasing consensus among philosophers of science that counterfactuals abound in science, especially when one reasons about scientific models (Dohrn, 2023; Iranzo-Ribera, 2022; McLoone, 2021; Tan, 2019). More precisely, many authors think that model-based reasoning is crucial to scientific practice[1] and that model-based reasoning is crucially related to counterfactual reasoning. Indeed, very often in science the theoretical claims we arrived at by reasoning about models are not intended to hold in the actual world, rather those theoretical claims are intended to hold in some highly idealized models (Shaffer, 2012). In this view, models can be regarded, in a sense to be specified, as worlds that are quite remote from the actual world.[2] Usually, possible worlds are understood to be possible in the sense that it is metaphysically possible for any one of them to have obtained (Ibidem). When we reason with models, we usually seek to establish whether a given statement *s* is true in a given model *M*. Counterfactuals are subjunctive conditionals whose antecedent is assumed to be false in the discourse (Goodman, 1947), i.e., false in the actual world, either because it did not occur or because it could not occur. The semantics for counterfactuals is thus apt to determine the truth value of statements that regard whether a given statement *s* holds in a model *M* that is actually false, since by assumption *M* irreducibly differs under some relevant respects from the actual world. So, one can conceive of the concept 'truth in a model' as truth "in a situation considered as counterfactual" (Dohrn, 2023, p. 1).

According to the standard semantics for counterfactuals (Lewis, 1973; Stalnaker, 1968), a counterfactual A > C (to be read as: *if it had been the case that A, then it would have been the case that C*) is true if there is some possible world, w_i, where the antecedent, A, and the consequent, C, are true that is more similar to the actual world, @, than any other possible world where A and the negation of the consequent, ¬C, are true. If, on the contrary, there is some possible world, w_j, where A and ¬C are true that is more similar to @ than any other possible world where A and C are true, the counterfactual is regarded as false. Finally, if there is no possible world where A is true, the counterfactual is said to be vacuously true.[3]

Counterfactuals with a metaphysically impossible antecedent, i.e., an antecedent that is true at no possible world, are called counterpossibles. The reason why, according to the standard semantics for counterfactuals, if there is no possible world where the antecedent is true, the counterfactual is said to be vacuously true, is that it is reasonable to require that any counterfactual whose consequent merely repeats

---

[1] The literature on model-based reasoning in science is quite vast. For an overview, see Magnani and Bertolotti (2017).

[2] It is usually thought that models cannot be equated with possible worlds, because possible worlds are usually thought to be consistent and complete (Menzel, 2023), while models need not to be consistent and complete (Iranzo-Ribera, 2022). Here we will assume that highly idealized models can be regarded as incomplete worlds (Shaffer, 2012), and that impossible worlds can be regarded as incomplete worlds. More on this below, sect. 6.

[3] It is worth noting that there are different opinions on how to make precise the idea that a possible world w_i is more similar to the actual world @ than another possible world w_j (see Starr, 2022, especially sects. 2.5 and 2.6, for an overview). We will not enter that debate here since it is beyond the scope of this chapter. Rather, in what follows we will assume that such a procedure to evaluate counterfactuals is viable.

its antecedent to be a necessary truth, i.e., a truth that holds in every possible world. Williamson (2018) clearly illustrates this point. Consider a counterfactual, A > C. The intension of A, |A|, is the set of possible worlds at which A is true. The intension of C, |C|, is the set of possible worlds at which C is true. The intension of the counterfactual, |A > C|, is a function of the intensions of the antecedent and the consequent:

(1) $|A > C| = f(|A|, |C|)$.

As regards |C|, what is relevant to determine the |A > C| is just the intersection of |A| with |C|, i.e., those possible worlds where the antecedent is true, and the consequent is also true. So, we have:

(2) $|A > C| = f(|A|, |A| \cap |C|)$.

Now, assume that A is true at no possible world, and so that |A| is the empty set, {}. This means that if the antecedent is true at no possible world, then the antecedent and the consequent are both true at no possible world. In this case we have:

(3) If $|A| = \{\}$ then $|A > C| = f(\{\}, \{\})$.

This means that all counterpossibles have the same intension, i.e., they are true at the same possible worlds, namely at no possible world. So, they are indiscriminate, in the sense that they could be regarded as all true or all false, but it is not possible to regard some counterpossibles as true and some other counterpossibles as false. If we introduce the apparently reasonable requirement that any counterfactual whose consequent merely repeats its antecedent to be a trivial necessary truth, i.e., true in any world that belongs to the set of all the possible worlds, W, we have:

(4) $|A > A| = W$.

If we take together (3), with C = A, and (4), it follows that $f(\{\}, \{\}) = W$. If we put $f(\{\}, \{\}) = W$ back into (3), we have:

(5) If $|A| = \{\}$ then $|A > C| = W$.

In other words, according to the standard semantics for counterfactuals, if the antecedent of a counterfactual is true at no possible world, then that counterfactual is true in every possible world. Counterpossibles are thus regarded as true solely because their antecedents are metaphysically impossible. Counterpossibles are said to be vacuously true. This is the so-called vacuity thesis in counterfactual semantics.

To better see this point, consider the following counterfactuals:

(6) If Hobbes had squared the circle, sick children in the mountains of South America at the time would have not cared. (Nolan, 1997, p. 544).

(7) If Hobbes had squared the circle, sick children in the mountains of South America at the time would have cared.[4]

That squaring the circle is impossible is a demonstrated mathematical truth. Mathematical truths are usually regarded as necessary truths (Yli-Vakkuri, Hawthorne, 2020). So, both (6) and (7) are counterpossibles since their antecedents are metaphysically impossible. Thus, according to the standard semantics for counterfactuals, both (6) and (7) are vacuously true. In this view, there is indeed no way to discriminate among counterpossibles, all counterpossibles are just uninformative, necessary truths.[5]

2. Counterpossibles and non-vacuism

Some philosophers think that this orthodox account of counterfactuals is unsatisfactory since there are cases in which we do discriminate counterpossibles. In their view, the semantic orthodoxy is unable to account for actual counterfactual reasoning, and so it should be amended. For instance, a far more intuitive way to evaluate (6) and (7) is to regard (6) as non-vacuously true and (7) as false (Berto *et al*., 2018). It seems reasonable to think that, even if Hobbes were able to square the circle and the sick children in the mountains of South America had known it, they had more urgent things to care about. Or consider the following countermetalogicals:

(8) If intuitionist logic were the correct logic, it would be impossible to validly infer *p* from ¬¬*p*. (Sandgren, Tanaka, 2019, p. 796).

(9) If intuitionist logic were the correct logic, it would be possible to validly infer *p* from ¬¬*p*.

If one assumes that the logic of the actual world is not intuitionist logic, (8) and (9) are counterpossibles, more precisely countermetalogicals. Counterfactuals of this kind "are commonplace in the contemporary study of logic": given "the proliferation of non-classical logics, it is often crucial to determine which inferences

[4] There are several ways in which the antecedent of a counterpossible can be impossible. For instance, in the antecedent of a counterpossible can figure: a mathematical impossibility, a logical impossibility, the negation of an identity, etc. Counterpossibles are then called accordingly: countermathematicals, counterlogicals, counteridenticals, etc. (Kocurek, 2021).

[5] That counterpossibles are uninformative if the vacuity thesis is accepted is the standard view. We will stick to that view. However, contrary to that view, see Dohrn (forthcoming), who claims that counterpossibles can still be informative despite being vacuously true.

would be valid if this or that logic were correct" (Ibidem). Since double negation elimination is invalid according to intuitionist logic, many authors regard (8) as non-vacuously true and (9) as false. So, the standard semantics for counterfactuals endorsed by many logicians and philosophers of logic seems not even able to adequately account for logicians' and philosophers' themselves reasoning practice.

The philosophers who are unconvinced by the standard semantics for counterfactuals reject the vacuity thesis and support the development of a non-vacuist semantics for counterfactuals. Other philosophers think instead that orthodoxy should be defended (Williamson, 2018; 2017a). For example, Williamson (2017a) thinks that if one responds in a "theoretically unreflective way" when is presented with statements like (6) and (7), one's response is "not always veridical" (p. 199). There are other cases in logic where our pre-theoretical inclination leads us astray. Williamson compares the vacuous truth of counterpossibles with the vacuous truth of empty universal generalizations. As an example, consider the following statements:

(10) Every golden mountain is a valley. (Williamson, 2017a, p. 205).

(11) Every golden mountain is a mountain. (Ibidem).

At first one might be tempted to regard (10) as false and (11) as true. But classical logic teaches us that, since there are no golden mountains, both (10) and (11) are vacuously true. The analogy between the case of counterfactuals and that of quantification is clear: the "impossibility of the antecedent corresponds to the emptiness of the subject term. For it is widely agreed that 'Every N Vs' is true if and only if the extension of N is a subset of the extension of V. Thus, as a special case, if the extension of N is empty, it is a subset of the extension of V, whatever V is, so 'Every N Vs' is true" (Ibidem). This means that our pre-theoretical assessment of the truth-value of a given statement is not always reliable. Sometimes a more reflective and theoretical approach is needed to provide the correct evaluation of a given statement. Turning to counterpossibles, the problem is that "processing a non-obvious counterpossible typically *feels* very like processing a non-counterpossible counterfactual" (Ibidem, p. 215). According to Williamson, what we do when we assess a given counterfactual A > C is that we suppose the antecedent, A, and see whether within the scope of A we accept the consequent, C. If we do, we regard the counterfactual A > C as true. If instead we do not accept C within the scope of A, we regard A > C as false. We tend to apply that suppositional procedure both when we assess counterfactuals with metaphysically possible antecedents and when we assess counterfactuals with metaphysically impossible antecedents, i.e., counterpossibles. This might be the reason why we tend to produce false judgments when it comes to counterpossibles. We might be unable to unreflexively appreciate the difference that the fact that the antecedent of a counterfactual is metaphysically impossible makes for assessing the truth-value of that counterfactual precisely in the same way in which we are usually unable to unreflexively appreciate the difference that the fact that the subject term is empty makes for assessing the truth-value of statements where it figures some universal quantification. So, according to this line of reasoning, that orthodoxy is unable to account for our intuitive assessment of counterpossibles should not be seen as a decisive reason to reject semantic orthodoxy about counterpossibles.

On the contrary, according to Williamson, there are strong theoretical reasons to defend orthodoxy. If one rejects the vacuity thesis, classical logic is in danger. Indeed, those who support the thesis that we should revise the standard semantics for counterfactuals, usually think that we should extend our semantics and admit of impossible worlds (more on this below). As Nolan states, there is a worry about impossible worlds: “the worry is that allowing such things respectability will bring the evils of nonclassical logic in their train” (Nolan, 1997, p. 543). The idea is that if one dismisses classical logic, one is unable to preserve and account for the rationality of science, and especially of mathematics, which is not just one of the sciences, but the science that is crucial for all other sciences. Williamson (2017b) defends the claim that classical logic is the one true logic by noting that classical logic is the logic on which the large part of mathematics is based. Since it would be unreasonable to dismiss such a huge and fundamental portion of mathematical knowledge, we should stick to classical logic.

Some non-vacuists object to such a defence of orthodoxy that strong support for non-vacuism comes precisely from the analysis of the role played by counterpossibles in scientific practice, even in mathematical practice (Tan, 2019; Jenny, 2018).

3. Counterpossibles and scientific practice

When in science we reason with models and regard a certain statement *s* as true in a model *M*, we reason counterfactually: we consider a given counterfactual, for instance: ‘if it were the case that *M*, it would be the case that *s*’, and regard it as true. As we have seen, this means that we think that the possible world w_i, which is as specified by *M* and where *s* is true, is more similar to the actual world @ than any other possible world that is as specified by *M* and where *s* is false. Counterfactual reasoning and model-based reasoning are deeply related. According to some philosophers, they are just one and the same. For instance, McLoone states that “a model can be formulated as a counterfactual conditional, where the antecedent of the conditional is a conjunction of all of the model’s assumptions, and the consequent is a conjunction of all of the model’s predictions” (McLoone, 2021, p. 12156). In this view, “scientific counterfactual reasoning is none other than model-based reasoning” (Iranzo-Ribera, p. 473).

According to several philosophers, in many cases some of the models one deals with in science appear to have assumptions that are metaphysically impossible, i.e., true at no possible world (Iranzo-Ribera, 2022; McLoone, 2021; Tan, 2019; Shaffer, 2012). This means that in science one has in many cases to deal with counterfactuals with a metaphysically impossible antecedent, i.e., counterpossibles. As noted, according to the standard semantics for counterfactuals, all counterpossibles are vacuously true. But many philosophers argue that scientific practice shows that counterpossibles are not always regarded as vacuously true by scientists. Indeed, in science at least some counterpossibles are regarded as non-vacuously true and others are regarded as false (Tan, 2019).

As an example, consider how water behaviour is modelled in hydrodynamics. Certain models of water represent it as a continuous, incompressible medium, so that the fluid flow can be modelled by means of the Navier-Stokes equations. In other words, those models "represent water *not* as being composed of discrete molecules but as if it were continuous" (Ibidem, p. 46). This is a clear case of idealization. The reason why scientists use such idealization is that it allows scientists to use mathematical tools to make calculations that provide accurate predictions of actual water's behaviour, predictions that would be impossible to make if such idealization were dropped. So, in this case scientists deal with the following counterfactual:

(12) If water were a continuous, incompressible medium, then it would behave as the Navier-Stokes equations describe. (Ibidem).

As Tan suggests, (12) can legitimately be regarded as a counterpossible. Indeed, there are reasons to think that its antecedent is metaphysically impossible. First of all, if one relies on standard, Kripkean way to deal with *a posteriori* necessity, i.e., necessary truths of which we acquire knowledge by means of experience, necessarily, water is identical to H_2O (Kripke, 1980). If necessarily water is identical to H_2O, i.e., water is made of discrete molecules in all possible worlds, it is metaphysically impossible for water to be a continuous and incompressible medium (Tan, 2019, p. 46). Secondly, if one is sceptical about Kripkean treatment of *a posteriori*, *de re*, necessary identity, it is in any case reasonable for one to consider the *de dicto* content of 'water', understood as a theoretical term, and claim that it is metaphysically impossible for the theoretical roles that water fills to be filled by any substance that is not composed of discrete particles. For example, nothing continuous and incompressible could act as a solvent or freeze into ice (Ibidem). That water is identical to H_2O can thus be regarded as a necessary truth.

So, (12) can be regarded as a counterpossible. This example is meant to support the claim that scientists deal with genuine counterpossibles when they deal with idealized models. The second point made by Tan (Ibidem) is that scientists do not regard all the counterpossibles they deal with as vacuously true. Suppose that scientists develop two distinct models of water behaviour, M_1 and M_2. Suppose also: 1) that both M_1 and M_2 assume that water is a continuous, incompressible medium; 2) that both M_1 and M_2 rely on the Navier-Stokes equations to model the fluid flow; 3) that they differ from each other with respect to some other parameter considered, for instance viscosity. M_1 and M_2 thus make different predictions. Suppose that scientists observe that actual water behaviour is better predicted by model M_1. So, scientists wish to accept M_1 and reject M_2. We can account for this situation in terms of counterfactuals. In this case, scientists consider the following counterfactuals:

(13) If water were a continuous, incompressible medium, then it would behave as M_1 predicts. (Ibidem, p. 47).

(14) If water were a continuous, incompressible medium, then it would behave as M_2 predicts. (Ibidem).

Now, (13) and (14) can be regarded as counterpossibles, since, for the reason just discussed, their antecedents can be regarded as metaphysically impossible. So, according to the standard semantics for counterfactuals, both (13) and (14) should be regarded as vacuously true. Nevertheless, (13) and (14) are not regarded as vacuously true by scientists. Rather, (13) is regarded as non-vacuously true, while (14) is regarded as false by scientists.

According to Tan (Ibidem) and other philosophers (see, e.g., Jenny, 2018; McLoone, 2021), this kind of examples shows that the philosophical orthodoxy about counterpossibles is inconsistent with scientific practice, and this fact provides a decisive reason for questioning the orthodoxy. Indeed, here we do not deal with the unreflective attitude that might lead us astray when we assess counterfactuals revolving around weird metaphysical hypotheses far remote from our daily experience. Here we deal with how scientists assess counterpossibles in their daily work. And their work provides us with scientific knowledge. If evaluation of counterpossibles plays a crucial role in scientific practice, and scientific practice produces genuine knowledge, we have to conclude that scientists are reliable when they assess counterpossibles. So, if scientists discriminate counterpossibles, orthodoxy is inconsistent with science. And this is a reason to reject orthodoxy (Tan, 2019). Indeed, philosophers of science usually defer to scientists when it comes to assess whether a given scientific practice should be regarded as acceptable.

4. An objection to counterpossibles in science

One might object that in science one does not really deal with counterpossibles, since what one deals with in science when one reasons counterfactually is at most some violations of laws of nature, and laws of nature are not to be regarded as metaphysically necessary. In this perspective, laws of nature are contingent, in the sense that they might have been different. So, when one deals with scientific models, one deals with counterfactuals with *physically* impossible antecedents, i.e., counterlegals, and not with metaphysically impossible antecedents. Thus, in science one does not really deal with counterpossibles.

There are two main rejoinders to this objection. First, let us admit, for argument's sake, that one can be sceptical of considering laws of nature as metaphysically necessary. However, mathematical truths are uncontroversially regarded as metaphysically necessary. And mathematics is usually regarded as one of the sciences. Counterfactual reasoning can be found even in mathematics (Jenny, 2018). In the context of mathematics, it is not contentious that if one deals with counterfactual reasoning, then one deals with counterpossibles. And the analysis of scientific practice in the field of mathematics seems to confirm that mathematicians discriminate counterpossibles. Jenny (2018), for instance, considers the case of relative computability. Relative computability is a central topic in computability theory (Dean, 2023, sect. 3.5). Computability theory studies what sets of natural numbers are algorithmically decidable. A set of natural numbers is algorithmically decidable if there is an algorithm that is able, in a finite number of steps, to solve the decision problem of that set, i.e., to decide for any natural number whether it is a member of that set. Gödelization, i.e., the technique introduced by Gödel (1931) in which a function assigns to each symbol and well-formed formula of some formal language a unique natural number, made possible to translate mathematical statements into natural numbers and represent mathematical problems as sets of natural numbers. For example, the VALIDITY PROBLEM is the set of natural numbers that encode the sentences of the predicate calculus that are logically valid (Jenny, 2018, p. 532). If the set of natural numbers that represents a given mathematical problem is algorithmically decidable, that problem is solvable by a procedure that can, at least in principle, be mechanized. If the set is not algorithmically decidable, that problem is not solvable by means of a mechanical procedure. So, to say that the VALIDITY PROBLEM is algorithmically decidable would be to say that there is an algorithm that would allow us to decide for any natural number that represents a sentence of the language of the predicate calculus whether it is a member of that set, and so whether it is logically valid (Ibidem). Turing (1936) demonstrated that the VALIDITY PROBLEM is not algorithmically decidable. It has also been demonstrated that other relevant sets of natural numbers are not algorithmically decidable, such as, for instance, the HALTING PROBLEM, "which encodes the problem of deciding whether a computer will eventually halt when it's given a certain input," and ARITHMETICAL TRUTH, "which encodes the true sentences of the language of arithmetic" (Jenny, 2018, p. 532).

Relative computability is the study of which sets of natural numbers would be algorithmically decidable if we assume that a given set of natural number is algorithmically decidable. The works of Turing (1939) and Post (1944) were fundamental for the development of relative computability. A central concept in relative computability is reducibility: a given set of natural numbers, B, is reducible to another given set of natural numbers, A, if the fact that A is algorithmically decidable implies that B is algorithmically decidable. More precisely, a set of natural numbers is algorithmically decidable if there is a Turing Machine (TM) that is able to compute it, i.e., to decide for any natural number whether it is a member of that set. Those sets that cannot be computed by a normal TM can be computed by an oracle Turing machine (OTM), a theoretical device introduced by Turing (1939). An oracle is a set of natural numbers that provides in a single computational step the correct answer for any question about whether a given natural number belongs to a given set of natural numbers that a normal TM is not able to compute. A given set of natural numbers, B, is Turing reducible to another given set of natural numbers, A, if there is an OTM that is able to decide for any natural number *n* whether *n* is a member of B when it has A as the oracle set. The concept of reducibility allowed researchers in relative computability to prove that not all problems are equally unsolvable. Some problems are more unsolvable than others. There is indeed a hierarchy of unsolvable problems. As Post writes, "the concept of reducibility leads to the concept of degree of unsolvability, two unsolvable problems being of the same *degree of unsolvability* if each is reducible to the other, one of lower degree of unsolvability than another if it is reducible to the other, but that other is not reducible to it, of incomparable degrees of unsolvability if neither is reducible to the other" (Post, 1944, p. 289).

Two results of relative computability are the following: the HALTING PROBLEM is reducible to the VALIDITY PROBLEM; the ARITHMETICAL TRUTH is not reducible to the VALIDITY PROBLEM. Now, consider the following counterfactuals:

(15) If the VALIDITY PROBLEM were algorithmically decidable, then the HALTING PROBLEM would also be algorithmically decidable. (Jenny, 2018, p. 533).

(16) If the VALIDITY PROBLEM were algorithmically decidable, then ARITHMETICAL TRUTH would also be algorithmically decidable. (Ibidem).

According to Jenny (2018), this is the kind of counterfactuals the mathematicians who work in relative computability usually deal with. It is reasonable to regard (15) and (16) as counterpossibles since their antecedents are in contrast with demonstrated mathematical truths. According to Jenny (Ibidem), it is also evident that mathematicians do not regard (15) and (16) as vacuously true. Given the results of computability theory, it is fair to say that mathematicians regard (15) as non-vacuously true and (16) as false. So, one either 1) denies that mathematics is a science, or 2) one has to admit *a*) that scientists deal with genuine counterpossibles and *b*) that scientists discriminate counterpossibles.[6]

A second rejoinder to the objection that in science one does not really deal with counterpossibles hinges on the role played by mathematics in explaining natural phenomena, an issue that received a lot of attention in recent years (Baker, 2009; Lange, 2013). Mathematical explanations of natural phenomena (MENP) are non-causal, scientific explanations in which an indispensable explanatory role is played

[6] One might deny that in mathematics one really deals with counterfactuals conditionals and claim that in mathematics one deals only with material conditionals. According to many philosophers, there is indeed no modal discourse in mathematics (see, e.g., Hodges, 2013), since modal discourse concerns alternative ways in which things can be, while mathematics is made of necessary truths. In this perspective, counterfactuals in mathematics are just material conditionals in disguise. To discuss this issue would bring us too far. It suffices to note that Jenny (2018, sect. 4) and Yli-Vakkuri and Hawthorne (2020, sect. 4.2) provide reasons as for why we should regard counterpossibles in mathematics as genuine counterpossibles. One of the most convincing reasons provided for thinking that there are genuine counterpossibles in mathematics is that counterfactuals conditionals in mathematics fail to conform to antecedent strengthening, which is valid for material conditionals (Jenny 2018, sect. 4).

by a mathematical fact (Baker, 2009). Since mathematical entities are usually regarded as non-causal, MENP are regarded as non-causal explanations. According to Lange (2013), MENP explain some worldly facts by showing how those facts are modally constrained by some mathematical fact. In this view, MENP explain in virtue of the extra modal force that mathematics has compared to the degree of necessity associated with ordinary physical laws. Indeed, mathematics is thought to be metaphysically necessary, while physical laws are just physically necessary. Those who support the idea that a theory of scientific explanation should be able to account in a unified manner for both causal and non-causal scientific explanations usually think that a counterfactual account of scientific explanation can make the case (Reutlinger, Colyvan, Krzyżanowska, 2022; Reutlinger, 2017). In this line of reasoning, MENP are cast in terms of counterfactuals. That the modal strength of mathematics constraints somehow the worldly fact that constitutes the *explanandum* of a MENP, and thus that explains why the *explanandum* is as it is by showing that it couldn't be in any other way, is shown by pointing out the counterfactual dependence of the *explanandum* on the mathematical fact that constitutes (at least an indispensable part of) the *explanans*. As an example, consider the following question: Why do hive-bee honeycombs have a hexagonal structure? Part of the explanation depends on evolutionary facts. But the explanation crucially rests on the mathematical fact that the hexagonal tiling is optimal with respect to dividing the plane into equal areas and minimizing the perimeter. This geometrical fact, known as the honeycomb conjecture, was proved by Hales (2001). In this view, the behaviour of hive-bees depends counterfactually on what is the optimal way to tesselate a surface into regions of greatest area with least total perimeter, i.e., a mathematical fact. The relevance of that mathematical fact in explaining hive-bees behaviour can be expressed by formulating the following counterpossible:

(17) If the optimal way to tesselate a surface into regions of greatest area with least total perimeter had not been via hexagons, hive-bees would not have built hexagonal honeycomb cells. (Baron, Colyvan, Ripley, 2017, p. 14).

Now consider the following counterpossible:

(18) If the optimal way to tesselate a surface into regions of greatest area with least total perimeter had not been via hexagons, hive-bees would still have built hexagonal honeycomb cells.

If one thinks that assuming a counterfactual theory of explanation is a promising way to account in a unified manner for both causal and non-causal explanations in science, and if one thinks that MENP are genuine scientific explanations, one has to admit that (17) and (18) are genuine counterpossibles, since the negation of a mathematical fact figures in their antecedents, and so that scientists deal with counterfactuals of that kind. Moreover, if one has to adequately explain hive-bees behaviour in terms of counterfactuals, one has to be able to regard (17) as non-vacuously true and (18) as false. So, MENP provide reason, albeit non-conclusive reason,[7] to think *a*) that scientists in their work deal with genuine counterpossibles and *b*) that scientists discriminate counterpossibles.

5. Counterpossibles and impossible worlds

To do justice of the use of counterpossibles in science, some authors think that we should adopt a non-standard semantics for counterfactuals, i.e., a semantics that allows for impossible worlds (McLoone, 2021). Impossible worlds can be characterized as ways things could not have been.[8] So, they are impossible in the sense that it is metaphysically impossible for any one of them to have obtained. By admitting of impossible worlds in one's semantics, one is able to evaluate a counterpossible as non-vacuously true or false.[9] In this view, a counterfactual A > C is true if there is some possible or impossible world, w_i, where the antecedent, A, and the consequent, C, are true that is more similar to the actual world, @, than any other possible or impossible world where A and the negation of the consequent, ¬C, are true, while it is false if there is some possible or impossible world w_j where A and ¬C are true that is more similar to @ than any other possible or impossible world where A and C are true. For instance, this non-vacuist semantics seems able to account for our intuitive assessment of the counterpossibles (13) and (14) that we discussed above:

(13) If water were a continuous, incompressible medium, then it would behave as M_1 predicts. (Tan, 2019, p. 47).

(14) If water were a continuous, incompressible medium, then it would behave as M_2 predicts. (Ibidem).

In the light of a non-standard semantic for counterfactuals, we regard (13) as non-vacuously true and (14) as false because we think that the impossible world w_i where water is a continuous, incompressible medium and behaves as predicted by

---

[7] This reason is non-conclusive because one might reject the claim that MENP are genuine scientific explanations, for instance because it is difficult to make MENP compatible with a naturalist stance. On this issue, see Sterpetti (2022).

[8] How one should conceive of impossible worlds is as a debated issue as the issue of how one should conceive of possible worlds. For an overview on how to conceive of impossible worlds, see Berto and Jago (2023), especially sect. 3.

[9] There are several difficulties that one has to address to make precise how the similarity relation between worlds has to be understood when one adds impossible worlds to the picture. This issue is relevant to determine how the worlds around @ are ordered, and so to assess counterfactuals. On this issue, see Berto and Jago (2023), Kocurek (2021) and Nolan (1997). This issue is beyond the scope of this chapter. In what follows we will assume that it is possible to evaluate counterfactuals by relying on some notion of similarity between worlds even when one adopts a non-standard semantics that admits of impossible worlds.

M_1, i.e., the model whose predictions have been empirically confirmed by scientists, is more similar to @ than any other impossible world where water is a continuous, incompressible medium and does not behave as predicted by M_1.

Consider now other two counterpossibles discussed above, namely (17) and (18), where another kind of metaphysical impossibility figures in the antecedents:

(17) If the optimal way to tesselate a surface into regions of greatest area with least total perimeter had not been via hexagons, hive-bees would not have built hexagonal honeycomb cells. (Baron, Colyvan, Ripley, 2017, p. 14).

(18) If the optimal way to tesselate a surface into regions of greatest area with least total perimeter had not been via hexagons, hive-bees would still have built hexagonal honeycomb cells.

In the light of a non-standard semantic for counterfactuals, we regard (17) as non-vacuously true and (18) as false because we think that the impossible world w_i where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees do not build hexagonal honeycomb cells is more similar to @ than any other impossible world where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees build hexagonal honeycomb cells.

Finally, consider again the two countermetalogicals that we discussed above, where still another kind of metaphysical impossibility figures in the antecedents:

(8) If intuitionist logic were the correct logic, it would be impossible to validly infer p from $\neg\neg p$. (Sandgren, Tanaka, 2019, p. 796).

(9) If intuitionist logic were the correct logic, it would be possible to validly infer p from $\neg\neg p$.

In the light of a non-standard semantic for counterfactuals, we regard (8) as non-vacuously true and (9) as false because we think: 1) that the actual world, @, is not an intuitionist world, since it is, for instance, a world where the correct logic is classical logic; and 2) that the impossible world w_i where the correct logic is the intuitionist logic and it is impossible to validly infer p from $\neg\neg p$ is more similar to @ than any other impossible world where the correct logic is the intuitionist logic and it is possible to validly infer p from $\neg\neg p$.

All those examples are genuine instances of counterpossibles that scientists ordinarily assess. Those examples also show that for the correct assessment of those counterpossibles a non-standard semantics for counterfactuals has to be adopted.

6. Impossible worlds and models

A minimal notion of ‘model’ is sufficient for accounting for model-based reasoning in terms of counterfactual reasoning. We will not enter ontological discussions about models.[10] In the context of a counterfactual understanding of models, models can be regarded as sets of propositions (Iranzo-Ribera, 2022). As already noted, it is easy to formulate models as counterfactuals (McLoone, 2021). Models can be regarded as constituted by assumptions, Ass., that, together with relevant information, R.I., entail some predictions, Pred. Assumptions can be regarded as the antecedent, A, of a counterfactual, A > C, while predictions can be regarded as its consequent, C. In this view, the predictions that constitute the consequent are the propositions that are entailed by the propositions that constitute the antecedent together with relevant information. So, a model *M* is equivalent to a counterfactual A > C where A = Ass. + R.I. and C = Pred., i.e.:

(19) M = (Ass. + R.I.) > Pred.

If we take the propositions that constitute *M* to be expressed by the sentences of a language $\mathscr{L}$, we can say that the set of sentences that constitute the antecedent of *M* is Π, while the set of sentences of $\mathscr{L}$ that constitute the consequent of *M* is Γ. A given model *M* can thus be seen as equivalent to the following counterfactual: Π > Γ, i.e.:

(20) $M = \Pi > \Gamma$.

This minimal notion of model is compatible with the idea of conceiving of models as incomplete worlds (Shaffer, 2012). Possible worlds, indeed, can be seen as maximally consistent and complete sets of propositions (Adams, 1974).[11] They are consistent in the sense that if those propositions are expressed by a set of sentences Σ in a language $\mathscr{L}$, Σ has a model (in the model-theoretic sense of the term), i.e., there is an interpretation of $\mathscr{L}$ under which every member of Σ is true. Σ is maximally consistent if it is consistent and there is no sentence *s* that can be added to Σ without Σ becoming inconsistent (Menzel, 2023). Possible worlds are complete in the sense of containing, for every proposition *p*, either *p* or its negation, ¬*p* (Ibidem). In this view, a proposition *p* is true if the sentence *s* that expresses it belongs to Σ. On the contrary, models in science need not to be necessarily

[10] For an overview on this issue, see Frigg and Hartmann (2023).

[11] For a different and more general characterization of possible worlds, see Priest (forthcoming). How the relation between possible worlds and propositions should be understood is a controversial issue. Indeed, there are several difficulties that arise if one identifies possible worlds with sets of propositions. Several difficulties arise also if one identifies propositions with sets of possible worlds. On this issue, see Bueno, Menzel and Zalta (2014). However, there is no room to properly deal with this issue here. Here, we adopt the view of possible worlds as sets of propositions, because this view is widespread in the literature devoted to the analysis of the relation between possible worlds and models, since it allows one to easily account for that relation.

consistent or complete. Indeed, a model usually does not allow one to determine, for every *p* whatsoever, whether it is the case that *p* or ¬*p*. And in fact, models are usually thought to be incomplete (Iranzo-Ribera, 2022). It might well be the case that, for a given *p*, in a model *M* there is neither *p*, nor ¬*p*. As Iranzo-Ribera writes, models "only enable the assignment of determinate truth-values to the […] propositions […] that fall within the scope of the model. Those that fall outside of a model's scope are indeterminate, for the model does not include the facts that would make these propositions true or false" (Iranzo-Ribera, 2022, p. 7). As an example, think again to models M_1 and M_2 that figure in (13) and (14) and that model water behaviour. If *p* = tomorrow there will be a solar eclipse, it is quite reasonable to think that in M_1 and M_2 one will find neither *p*, nor ¬*p*, and so that one will be unable to determine whether *p* is true.

Nor models in science need necessarily to be consistent or maximally consistent. First of all, there are useful models in science that are not, strictly speaking, consistent. Indeed, it is well known that models in science "are sometimes logically inconsistent, or at least are idealized in logically inconsistent ways" (Tan, 2021, p. 12). As for examples, consider "Bohr's theory of the atom", the "old quantum theory of blackbody radiation", "Kirchhoff's diffraction theory", and "Lorentz's early theory of the electron" (Ibidem). In all those cases, "it is possible to derive both *p* and ¬*p* within the theory for some *p*" (Ibidem). Bohr's theory of the atom, for instance, "ends up representing orbiting electrons both as emitting radiation and not emitting radiation" (Ibidem). This might be due to the fact that some assumptions within that theory "require certain quantum mechanical principles to be true, while other assumptions require contradictory classical principles to be true" (Ibidem). This is just an example of a phenomenon on which there is a large consensus in the literature, namely that it is possible that a scientific theory or model that is useful, i.e., it is able to provide correct predictions, taken as a whole contains, or allows the derivation of, some contradiction (Vickers, 2013; Tan, 2021).

Secondly, models need not to be maximally consistent. This is transparent in all those cases in which adding more relevant details or variables to a model, albeit possible, would not be regarded as useful by scientists, since it would not increase the model's capacity of providing some understanding of the modelled phenomenon (Elgin, 2007). As an example, consider abstract models of traffic jam. Indeed, one can either develop models which mimic the causal effects of the cars on one another or develop models which abstract away from any of the causal relationships found in the previous kind of models. For instance, we might "imagine a representation of a given system of cars that simply represented what the system would look like after all the cars had achieved some fixed velocity for a given density of cars" (Pincock, 2012, p. 5). It is usually the case that adding some other features to a model of that kind would not render the model inconsistent, even if those added features would probably not render it more effective either. This means that in the case of many models in science, some *s* could be added to Σ without making it inconsistent. So, models in science need not to be maximally consistent.

7. Counterpossibles and highly idealized models

Someone might contest the idea that, when we look at scientific models in terms of counterfactuals, models should necessarily be assessed by considering impossible worlds for the sole reason that models might be either incomplete, inconsistent, or both. One might claim that what is really necessary for a counterfactual to be regarded as a counterpossible is that the antecedent of that counterfactual be metaphysically impossible. To address this issue, we will restrict our attention to highly idealized models (Shaffer, 2012), i.e., models among whose assumptions figure some *radical* idealization (Rice, 2021). Indeed, radical idealizations are such that make the antecedent of a model, understood as a counterfactual, not just false because it differs to some extent from what is actual of the model's target. Rather, radical idealizations are such that make the antecedent of a model, understood as a counterfactual, metaphysically impossible. Thus, a counterfactual where a radical idealization figures in the antecedent can be regarded as a genuine counterpossible. As an example, think again to how water is modelled in hydrodynamics. When one considers water as a continuous, incompressible medium one is adopting a radical idealization, since the distortion of the model's target is such that the highly idealized model cannot be regarded as an approximate representation of the model's target in any relevant sense (Ibidem). Nor are highly idealized model intended to provide an accurate representation of the model's target. Rather, they are thought to be able to point out the counterfactual dependence of the *explanandum* on the *explanans*. According to Rice, a model "can be used to show that there is a relationship of counterfactual dependence […] between [the *explanans*] and [the *explanandum*], even if it drastically distorts many aspects of [the *explanans*] and [the *explanandum*]," and even if it "pervasively misrepresents the processes or mechanisms that connect [the *explanans*] with [the *explanandum*]" (Ibidem, p. 147).

Does the fact that we restrict our attention to highly idealized models severely restrict the scope of our argument, and so its relevance? The answer seems to be in the negative, since highly idealized models are ubiquitous in science (Rice, 2021; Shaffer, 2012). Rice (2021), for instance, convincingly argue that radical idealizations in science are indispensable, ineliminable, and pervasive. Radical idealizations are indispensable, since these idealizations allow mathematical tools be applicable to account for the model's target. So, radical idealizations indispensably contribute to the model's capacity of providing predictions or explanations. Think again to how water is modelled in hydrodynamics. If scientists would not regard water as a continuous, incompressible medium, they would be unable to apply the Navier-Stokes equations to water behaviour. Radical idealizations are also ineliminable, because it is often the case that if one removes radical idealizations to replace them with more accurate representation of the target's models, the model becomes mathematically intractable, for instance because it exceeds our computational capacity. Finally, radical idealizations are pervasive, since modern science strongly relies on mathematics to account for investigated phenomena in a rigorous way, and to make the relevant mathematics applicable to worldly phenomena very often some radical idealization is required.

The specification of the models we are considering should make easier to agree upon the following claims: 1) highly idealized models in science should be understood in terms of counterpossibles; 2) to adequately account for scientists' evaluation of those models we should adopt a non-standard semantics for counterfactuals that admits of impossible worlds;[12] 3) highly idealized models can be regarded as impossible worlds, since they are equivalent to incomplete worlds that represent metaphysically impossible scenarios.

That highly idealized models can be regarded as *incomplete* worlds is well illustrated by Shaffer (2012), who accounts for the making of highly idealized models in science in terms of a process by which, starting from a base model M which models a given phenomenon p, a new, idealized model M' is derived by a series of successive simplifications and distortions of the features that relate M to p. So, even if one remains unconvinced that models in general are incomplete, it is clear that highly idealized models cannot be complete, since, by definition, they are constructed by subtracting elements, i.e., propositions, from base models. That highly idealized models are incomplete worlds means that highly idealized models are more similar to impossible worlds than to possible worlds. Indeed, impossible worlds can be seen as inconsistent and incomplete sets of propositions, while possible worlds, as already noted, are usually regarded as complete and consistent sets of propositions.[13]

Moreover, that highly idealized models can be regarded as *impossible* worlds, since they are incomplete worlds that depict metaphysically impossible scenarios, is well illustrated by Rice (2021), who, as already noted, accounts for how the distortion of relevant features of the modelled phenomenon introduced by radical idealization is such that makes the scenario depicted by a highly idealized model metaphysically impossible.

---

[12] It is worth noting, even if in passing, that although the great majority of those who support non-vacuism supports the extension of the standard semantics for counterfactuals to impossible worlds, there are some exceptions. For instance, French, Girard and Ripley (2022) argue that non-vacuism about counterpossibles can be reconciled with classical logic. Here we adhere to the mainstream view on the issue, namely that the adoption of some non-classical logic is the best choice if one adopts non-vacuism. We do not enter the debate on which kind of non-classical logic should be adopted if one adopts non-vacuism, because that issue is irrelevant for the aim of this chapter.

[13] This issue is debated. Cf. Nolan (2013, p. 368): "Whether impossible worlds should be complete is a trickier matter. Some are not prepared to call something a world unless it is complete, so some of the disputes may turn out to be terminological ones about which impossibilities are impossible worlds and which are mere impossibilities. However, 'worlds' or not, such incomplete impossibilities may play a significant world-like role in theories of conditionals." Here we assume that incomplete worlds are admissible, and so that impossible worlds can be regarded as incomplete worlds.

Now, if models need not be (maximally) consistent and complete, “possible-worlds-based semantics for counterfactuals do not naturally apply to scientific models” (Iranzo-Ribera, 2022, p. 7). The most widespread reaction to this fact has been to support the idea that one should extend the standard semantics for counterfactuals with impossible worlds (McLoone, 2021). In this view, models can be regarded as impossible worlds, and any assessment of whether a given sentence s is true in a model M can be regarded as an assessment of a counterpossible $\Pi > \Gamma$, with $s \in \Gamma$.

To recapitulate, if model-based reasoning is equivalent to counterfactual reasoning, and highly idealized models can be regarded as impossible worlds, and highly idealized models are ubiquitous in science, it is reasonable to conclude that to account for how we reason with models in science we need to adopt a non-standard semantics for counterfactuals that admits of impossible worlds. More precisely, we can say that a highly idealized model M can be regarded as equivalent to a counterpossible $\Pi > \Gamma$, and that $\Pi > \Gamma$ is true if there is some impossible world, w_i, where the sentences in Π are true and the sentences in Γ are also true that is more similar to the actual world, @, than any other impossible world where the sentences in Π are true and the sentences in Γ are not true, while it is false if there is some impossible world w_j where the sentences in Π are true and the sentences in Γ are not true that is more similar to @ than any other impossible world where the sentences in Π are true and the sentences in Γ are also true.

8. Scientific realism and modal realism

The ubiquity of counterfactual reasoning in science raises several issues if one wishes to take a realist stance on science. For instance, if one considers that the main thesis of scientific realism is that our best scientific theories are true or approximately true (Chakravartty, 2017); and that scientific realists usually commit themselves to the conception of truth as correspondence or to some variant of it (Sankey, 2008), where the correspondence relation is intended to hold between the language in which the theories are formulated and the mind-independent world, it is easy for one to see why it is not easy for a scientific realist to provide a counterfactual account of the role played by radical idealization in science. Radical idealizations make scientific models to irreducibly *differ* from the actual phenomena those models are intended to model, to the extent that it is not even possible to regard those models as approximately true representations of those phenomena (Rice, 2021). And indeed, some philosophers have argued that idealizations cause problems for scientific realism (see e.g., Odenbaugh, 2011). Nevertheless, some other philosophers think that it is possible for one to maintain scientific realism even if one acknowledges the relevance of counterfactual reasoning in science and the pervasiveness of highly idealized models (Shaffer, 2012; Rice, 2021). The price to pay is that one has to adopt a less straightforward conception of realism about science. In a nutshell, the idea is that realism about science should not be interpreted as the claim that scientific theories and models provide us with some (approximately) true representation of the world. Rather, scientific realism should be understood as characterized by two claims, namely 1) the claim that it is possible to assign a definite truth value to any theoretical claim that is intended to hold in some highly idealized model (Shaffer, 2012); and 2) the claim that highly idealized models are able to provide us with true modal information about models' targets, despite those models do not provide any accurate representation of the actual state of models' targets (Rice, 2021; Shaffer, 2012). So, what we should be realist about are not theories or models, but the modal information provided by theories and models.

As regard 1), it implies that one should reject the vacuity thesis in order to be able to discriminate counterpossibles, and so to assign a definite truth value to any theoretical claim that is intended to hold in some highly idealized models.

As regard 2), it has two main implications. First, it implies that one adopts a counterfactual account of scientific explanation in order to account in a unified manner for both causal and non-causal explanations, since highly idealized models are usually unable to provide causal explanations, because they do not accurately reflect the processes or mechanisms that occur in the target phenomenon, rather they usually provide some form of non-causal explanation by pointing out the relationship of counterfactual dependence that obtains between the *explanans* and the *explanandum* by means of those mathematical tools that idealizations make possible to apply to the target phenomenon. This makes the explanations provided by highly idealized models similar the MENP discussed above (sect. 4). As in the case of the explanation of why hive-bees construct hexagonal honeycombs, where the counterfactual dependence of honeycombs' shape on a mathematical fact allows one to clarify what comb construction abilities were metaphysically possible or impossible for hive-bees to acquire during their evolutionary path independently of any information about the contingent evolutionary path hive-bees actually took, highly idealized models provide us with genuine knowledge about what is metaphysically possible or necessary of model's target independently of any accurate representation of actual features of model's target (Rice, 2021). Moreover, a counterfactual account of scientific explanation is able to account for the relation that obtains between models and explanations in a way that is consistent with the account of the relations that obtains between models and possible worlds outlined above (sect. 6). In this view: *a*) possible worlds are complete and consistent sets of propositions; *b*) models are not necessarily complete and consistent sets of propositions; and *c*) an explanation is a set of true propositions sufficient to explain the target *explanandum*. So, a model M of ϕ is also an explanation of ϕ only if it includes a set of true propositions, E, sufficient to explain ϕ (Rohwer, Rice, 2016). If one deals with a highly idealized model, and highly idealized models are understood in terms of counterpossibles, i.e., if $M = \Pi > \Gamma$, M is an explanation of ϕ if one regards $\Pi > \Gamma$ as non-vacuously true and $E \subset \Gamma$.

Second, 2), i.e., the claim that highly idealized models are able to provide us with true modal information about models' targets, despite those models do not provide any accurate representation of the actual state of models' targets, implies that one commits oneself to some form of modal realism, since in this view the only thing left to the scientific realist to be realist about is the modal information that one extracts from highly idealized models by assessing counterpossibles. If one wishes to take modal claims at face value to maintain the standard realist attitude towards the truth, i.e., the idea that a proposition is true because there exists something that makes it true, it seems that the only route that one can take is to adopt some form of modal realism. Modal realism is the view according to which there exists possible worlds. There are several forms of modal realism (Menzel, 2023; Berto, Jago, 2023). For instance, according to *genuine modal realism*, mainly advocated by Lewis (1986), there exist other concrete worlds beyond the actual one, so modal claims are made true by what exists in those possible worlds. According to abstractionism, instead, possible worlds do exist but are non-concrete objects. For instance, some abstractionist philosophers think of possible worlds as sets of propositions and regard propositions as abstract entities. Moreover, since according to the variant of scientific realism that we are considering one is able to extract *true* modal information from models that do not accurately represent the model's target by means of assessing counterfactuals in whose antecedents figure some *metaphysical* impossibility, and that to do that one has to adopt a non-standard semantics for counterfactuals that admits of impossible worlds, it seems that one has to adopt a realist stance not just on possible worlds, but also on impossible worlds. According to many philosophers, indeed, there is "absolutely no cogent (in particular, non-question-begging) reason to suppose that there is an *ontological* difference between merely possible and impossible worlds" (Priest, 1997, p. 581). As there are several forms of realism about possible worlds there are several forms of realism about impossible worlds, although realism about impossible worlds is more controversial (for an overview, see Nolan, 2021; Berto, Jago, 2023, especially sect. 3). For instance, according to the *extended modal realism* advocated mainly by Yagisawa (1988; 2010), we have to conceive of impossible worlds in the same way in which we conceive of possible worlds. In this view, impossible worlds are concrete worlds where impossibilities are instantiated. Abstractionists can analogously extend their favourite kind of abstractionism about possible worlds to impossible worlds. For instance, if one conceives of possible worlds as maximally consistent sets of propositions, one can conceive of impossible worlds as sets of propositions that are inconsistent and/or incomplete. We will not enter the debate about what form of realism about possible and impossible worlds should be preferred, since our argument does not relevantly depend on what form of realism about possible and impossible worlds one commits oneself to. What is relevant to stress is that modal realism, understood as realism about possible and impossible worlds, is the only view that allows a straightforward realist treatment of modal notions in the context of a realist account of the role played by highly idealized models in science.

9. Scientific realism and the epistemology of modality

In the rest of this chapter, I will focus on just one of the objections that can be raised against the variant of scientific realism defended by Shaffer (2012) and Rice (2021) by elaborating on some arguments developed by Rescher (2020; 2019; 2009; 2006). As seen in the previous section, it seems that such variant of scientific realism should be combined with some form of modal realism. Now, realism about modality, as any form of realism, implies that there are modal facts that are mind-independent (Williamson, 2007) and that the truths about those modal facts are counterfactually independent of any epistemic agent. As an example of that kind of independence, consider realism about mathematics. In this view, mathematical truths are counterfactually independent of any epistemic agent. Indeed, since mathematical truths are regarded as metaphysically necessary, the following counterfactual can be formulated:

(21) Had there been no intelligent life, [mathematical] truths would still have remained the same. (Linnebo, 2017, p. 9).

If one regards (21) as true it means that one adopts a realist stance on mathematics. The same occurs in the case of modality. Consider the following counterfactual:

(22) Had there been no intelligent life, modal truths would still have remained the same.

If one regards (22) as true it means that one adopts a realist stance on modality, namely that one regards modal facts and truths as mind-independent. Consider now the issue of how one can acquire knowledge about the modal domain. It is usually thought that our ability in dealing with counterfactuals is crucial to provide an adequate epistemology of modality (Clarke-Doane, 2019; Kroedel, 2012; Williamson 2007). In other words, in this view, we acquire modal knowledge by reasoning counterfactually. This means that we can acquire knowledge about mind-independent truths by means of some epistemic, i.e., mind-dependent, capacity, such as the ability to conceive of alternative scenarios and to counterfactually reason about those scenarios.[14]

As an example, it is usually held that we can reliably conclude that it is metaphysically possible that p, $\Diamond p$, if and only if we consider p and find that no contradiction, $\bot$, follows from p. More precisely, we consider the following counterfactual:

(23) $p > \bot$,

[14] There are several proposals that one can find in the literature on how an adequate epistemology of modality should look like. For an overview, see Mallozzi, Vaidya and Wallner (2023). Details about the different theories in the epistemology of modality are not relevant to our purposes.

i.e., "had it been the case that *p*, then a contradiction would have obtained", and regard it as false, i.e., we judge that it is not the case that a contradiction would be the case if *p* were the case, $\neg(p > \bot)$, because we are able to conceive of a scenario in which *p* is the case and to develop such scenario in order to see whether a contradiction can be derived as a consequence of assuming that *p* is the case (Williamson, 2007). So, in this view there is such a close logical relationship between the counterfactual conditional and the metaphysical modalities that the following equivalence is thought to hold:

(24) $\Diamond p$ if and only if $\neg(p > \bot)$. (Kroedel, 2012, p. 5).

Keep in mind that we are dealing with metaphysical possibility, i.e., we are considering only possible worlds. In this restricted context, one can reason as follows. Given the standard semantics for counterfactuals, (23) is false if there is some possible world where the antecedent, *p*, and the negation of the consequent, $\neg\bot$, are true that is more similar to the actual world, @, than any other possible world where *p* and $\bot$ are true. Since, by assumption, there is no possible world where a contradiction is true, given that possible worlds have to be complete and *consistent*, the falsity of (23) implies the possibility of *p* precisely because (23) can only be false if there is some possible world in which *p* is true (Clarke-Doane, 2019, p. 269). If *p* were true at no possible world, (23) would be vacuously true, not false.

If we have to be realist about possible and impossible worlds because we aim to maintain a realist stance on science as Shaffer (2012) and Rice (2021) suggest, our capacity of assessing counterpossibles need to be reliable. As noted, it is by counterfactually reasoning that we can assess what is metaphysically possible or necessary. And it is by assessing what it is metaphysically possible or necessary that we assess whether a given counterpossible is true or false, even in science.

More precisely, in order to compare two distinct possible or impossible worlds, say w_i and w_j, and conclude that w_i is closer to the actual world, @, than w_j, we have to be reliable when we estimate the ordering of possible and impossible worlds. We have so far assumed (see above, fn. 3 and 9), at least for the sake of the argument, that there is a unique and uncontroversial way to order possible and impossible worlds. Let's maintain that assumption. What is crucial for the reliability of our assessment of counterpossibles is that we have to be able to consider, at least in principle, *all* the possible and impossible worlds that obtain. Only if we are able to exhaust the space of metaphysical possibilities and impossibilities, we can fairly conclude that, for instance, there is *no other* impossible world that is closer to @ than w_i and regard our conclusion about a given counterpossible as genuine modal *knowledge*. Otherwise, it might well be the case that one of the non-examined worlds be closer to @ than w_i. As an example, consider again the hive-bee counterpossible:

(17) If the optimal way to tesselate a surface into regions of greatest area with least total perimeter had not been via hexagons, hive-bees would not have built hexagonal honeycomb cells. (Baron, Colyvan, Ripley, 2017, p. 14).

As previously noted, we regard (17) as non-vacuously true because we think that the impossible world w_i where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees do not build hexagonal honeycomb cells is more similar to @ than any other impossible world where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees build hexagonal honeycomb cells. This assessment can be understood as equivalent to the claim that all impossible worlds have been taken into consideration (or, at least, to the claim that it could be possible, at least in principle, to consider all impossible worlds) and no world where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees build hexagonal honeycomb cells has been found to be more similar to @ than those impossible worlds where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees do not build hexagonal honeycomb cells. What if not all the impossible worlds that obtain are taken into consideration? If some world is not examined, how can one exclude the possibility that there might be an impossible world where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees build hexagonal honeycomb cells that is more similar to @ than those impossible worlds where the optimal way to tesselate a surface into regions of greatest area with least total perimeter is not via hexagons and hive-bees do not build hexagonal honeycomb cells? After all we are dealing with *impossible* worlds, and impossible worlds might be quite surprising (Berto, Jago, 2023). And indeed, the way in which impossible worlds might be ordered with respect to @ is a controversial matter (Ibidem).[15] It thus seems that it is only by considering all possible and impossible worlds, and by performing a sort of eliminative reasoning, that one can be reliable when it comes to assessing the similarity to @ of worlds. So, if we have reason to suspect that we are not able, not even in principle, to consider and examine all metaphysical possibilities and impossibilities, i.e., all possible and impossible worlds, our assessment of any given counterpossible would be just a tentative assessment, a hypothesis relatively confirmed, i.e., confirmed by the possibilities and impossibilities that we have been able to consider, but not absolutely confirmed,

---

[15] For instance, not all philosophers are convinced by the idea that any possible world is more similar to the actual world, @, than any impossible world (see, e.g., Nolan, 1997; Vander Laan, 2004). This is the so-called Strangeness of Impossibility Condition (Nolan, 1997, p. 550). According to this condition, possible worlds form a sphere around @ and so are more similar to @ than any impossible world. Some philosophers object that some slightly deviant impossible worlds, such as those impossible worlds where just a local inconsistency holds, might be more similar to @ than some possible but very weird worlds, such as those possible worlds where the laws of physics are very different from the laws that hold at @. On this issue, see Berto and Jago (2023, especially sect. 4.2).

i.e., confirmed by an examination of *all* possibilities and impossibilities.[16] But "relative confirmation has no established connection to truth-likeness, even on the assumption that absolute confirmation (in some non-subjective sense) does indicate truth-likeness" (Rowbottom, 2019, p. 3949). This means that if there are reasons to think that we are not able to consider all metaphysical possibilities and impossibilities we should refrain to regard our assessment of counterpossibles as able to provide us with genuine modal knowledge. And if this is the case, the variant of scientific realism proposed by Shaffer (2012) and Rice (2021) might be in trouble. Knowledge indeed requires truth. If what we should be realist about is just the modal information provided by highly idealized models, whether we are able to acquire genuine modal knowledge depends entirely on our ability in making correct assessment of counterpossibles.[17]

10. Scientific realism and inconceivable possibilities

So, the question is: Do we have any reason to suspect that we are not able to consider and examine all metaphysical possibilities and impossibilities when we assess counterpossibles in science? In this section, I suggest that some arguments, originally developed by Rescher (2020; 2019; 2009; 2006) to show that there might be unknowable facts, provide us with sufficient theoretical reason to suspect that we are not able to consider all the metaphysical possibilities and impossibilities when we assess counterpossibles.

In a nutshell, Rescher's arguments revolve around the idea that we can acquire knowledge about mind-independent facts only by means of some epistemic, i.e., mind-dependent, capacity, such as, for instance, conceivability, and that our epistemic capacities cannot but be related to our linguistic capacity, for "we humans have to conduct our conceptualizing business by means of language" (Rescher, 2020, p. 33). Our capacity of acquiring knowledge about the world is thus limited

[16] The point is not that our assessment might be wrong. Obviously, error might always occur. We are not unfairly requiring infallibility on the part of the modal realist. The point is whether there is some principled reason for doubting that our assessment can be reliable.

[17] It is worth noting, albeit in passing, that for the elaboration of my argument, besides the debt I owe the work of Rescher, I also clearly owe a debt to the work of Stanford (2006), who developed his New Induction against scientific realism by pointing out the problem of the unconceived alternatives. For this reason, I also indirectly owe a debt to the work of Sklar (1981) and van Fraassen (1989), who are among Stanford's forerunners, especially for the criticism they raised against the inference to the best explanation. The core of all those arguments against realism is basically the same: if one wishes to adopt a realist stance on *p* by means of some kind of eliminative reasoning, one needs to demonstrate to be able to consider all the possible alternatives to *p*. If there are reason for thinking that one is not able to consider all the possible alternatives to *p*, then there are reason for thinking that eliminative reasoning is not able to support a realist stance on *p*. All those arguments provide reason for thinking that one is not able to consider all the possible alternatives to *p*.

by the limits of language. Indeed, "linguistic formulation is a recursive process—exfoliating claims from a finite vocabulary via finite grammatical principles. And this means that we can realize at most a" countable number of statements (Ibidem, pp. 33-34). Since the set of all the truths that we are able to express is a subset of the set of all the statements that we are able to realize, it follows that we can express at most a countable number of truths (Rescher, 2006, p. 45). But, if one adopts a realist stance on facts and metaphysical possibilities, i.e., if one thinks that facts and metaphysical possibilities are mind-independent, there is no reason to think that facts and metaphysical possibilities are limited in the same way.

Assume, for the sake of the argument, that conceivability is a guide to metaphysical possibility.[18] If one takes a realist attitude towards conceivable possibilities, one conceives of them as mind-independent entities. Better, from a realist point of view, conceivable possibilities, *qua* metaphysical possibilities, are independent from what happens in the world, i.e., those possibilities would be conceivable independently of what were the case: the "domain of possibility […] is independent of and detached from what actually happens in the world" (Rescher, 2020, p. 34). As said, conceivability *qua* epistemic capacity cannot be decoupled by our linguistic capacity, and so is limited by the limits of language. This means that it might be the case that we are able to conceive at most a countable number of metaphysical possibilities, while metaphysical possibilities are in fact uncountable, i.e., they exceed conceivable possibilities. So, there might be metaphysical possibilities that we are not even able to conceive, i.e., inconceivable possibilities.

To claim that there are inconceivable possibilities amounts to claim that there are possible worlds that we are not able to take into consideration when we assess counterfactuals. Moreover, in modal logic, that it is possible that p, $\Diamond p$, means that there is some possible world where p holds, while that it is necessary that p, $\Box p$, means that p holds at every possible world. When only possible worlds are considered, $\Diamond p$ and $\Box p$ are so inter-definable: it is possible that p is equivalent to the negation of the necessity of the negation that p, i.e., $\Diamond p = \neg\Box\neg p$. One is tempted to claim that if there are reasons to think that there might be some possible world that we cannot have knowledge of and at which a metaphysical and inconceivable possibility, say ψ, holds, one has also reasons to think that there might also be some impossible world that we cannot have knowledge of, namely the impossible world where that metaphysical possibility is denied, a world where the negation of ψ is necessary, i.e., where $\Box\neg\psi$ holds. So, if there are inconceivable possibilities, it seems that there might also be inconceivable impossibilities.[19] But when impossible worlds are considered, things get more nuanced.[20] Some philosophers think that, in general, there should be no restriction on how impossible an impossible world can be (Nolan, 1997; Zalta, 1997). For example, according to Nolan, "for every

[18] On this issue, see Yablo (1993).

[19] It is perhaps worth noting that I'm not claiming that this is the only way to support the claim that there might be inconceivable impossibilities. There might well be other ways to support that claim, I'm just limiting my analysis to this route here.

[20] For an overview, see Berto and Jago (2023, especially sect. 5).

proposition which cannot be true, there is an impossible world where that proposition is true" (Nolan, 1997, p. 542). So, one might think that, despite how weird it might seem to one, there might well be an impossible world w_j at which $\Box\neg\psi$is true. That world seems weird because it is hard to make sense of the claim that something is necessary in just one world. That $\Box\neg\psi$ is true just in w_j is in contrast with the standard way in which necessity is defined. And indeed, according to many semantics that admit of impossible worlds, possibility and necessity should be defined only with respect to possible worlds (Berto, Jago, 2023). One might simply reply that since w_j is an *impossible* world, it is normal that weird things might happen in it. However, ψ might well be true at more than just a world. What is relevant is that all the possible worlds at which ψ is true are worlds that we cannot have knowledge of. In this case, there might be several impossible worlds that we cannot have knowledge of and at which $\Box\neg\psi$ is true. In this case, it might be said that $\neg\psi$ is necessary at w_j if $\neg\psi$ is true at all the impossible worlds that are accessible from w_j. The accessibility relation might here be construed just in terms of similarity between worlds.[21]

However, how to account for modal notions in a semantics that admits of impossible worlds is a controversial issue. For instance, in non-normal semantics provided by Kripke (1965) for systems of modal logic S2 and S3, all $\Box$-formulas are regarded as false. According to those semantics, one could not claim that $\Box\neg\psi$ might be true at some impossible world if ψ is true at some possible world. But there are other systems, such as S0.5, a modal system introduced by Lemmon (1957), whose semantics, firstly developed by Cresswell (1967) includes non-normal worlds at which formulas that begin with a modal operator such as $\Box$ are assigned arbitrary truth values. In that semantics, complex formulas are treated as atomic. So, according to that semantics, $\Box\neg\psi$ might well be true at some impossible worlds. However, it has also to be said that "this setting makes the inter-definability of $\Diamond$ and $\Box$ via negation fail" (Berto, Jago, 2019, p. 101). In this case, the derivability of the existence of an impossible world we cannot have knowledge of from the existence of a possible world we cannot have knowledge of would be less straightforward.

This brief discussion about inconceivable impossibilities is not intended to provide a conclusive argument for the existence of inconceivable impossibilities, it just aims to show that it might be the case that, if there are inconceivable possibilities, there might also be inconceivable impossibilities, and so that there might be possible and impossible worlds of which we cannot have knowledge. If this is the case, there would be a principled reason to doubt that our assessment of counterpossibles is reliable.

Now, examples of inconceivable possibilities cannot obviously be provided. So, why should we think that metaphysical possibilities outnumber conceivable possibilities, and so that it is reasonable to expect that there are inconceivable

---

[21] Cf. Kiourti (2010, p. 136): "a world w_1 is accessible from another world w_0 under the accessibility relation R^M if and only if w_1 is similar to w_0 with respect to a set of base facts M about w_0."

possibilities? Arguments originally provided by Rescher (2009; 2006) for showing that facts are uncountable while truths are countable, and so that there are unknowable facts, can be extended to cover the case of conceivable possibilities and show that there are inconceivable possibilities since conceivable possibilities outnumber the possibilities that are effectively conceivable. And in fact, those argument have been extended by Rescher himself along that line (Rescher, 2020; 2019). Moreover, metaphysical possibilities and impossibilities, from a realist point of view, are mind-independent, they are modal facts. If worldly facts, i.e., actual aspects of the world's state of affairs, are uncountable, *a fortiori* modal facts are uncountable, since usually worldly facts are regarded as a subset of modal facts. So, Rescher's arguments, to the extent that they are convincing, constitute a crucial element for developing a genuine challenge to the variant of scientific realism that is at stake.

The first argument to support the claim that facts are uncountable is the following *reductio ad absurdum*. Assume that the totality of facts forms a countable set, F. By Cantor's theorem, the cardinality of the power set of F, i.e., the set of all the subsets of F, $\mathcal{P}(F)$, is strictly greater than the cardinality of F. But every member of $\mathcal{P}(F)$ is itself a fact, since any complex of facts is still a fact, and so it should be possible to put every member of $\mathcal{P}(F)$ in correspondence with some unique member of F. But this means that the cardinality of $\mathcal{P}(F)$ cannot be strictly greater than the cardinality of F. Contradiction. Thus, the totality of facts does not form a countable set. (Rescher, 2009, p. 58).

The second argument for supporting the claim that facts are uncountable is as follows. Suppose that there is a list, F, that is a complete enumeration of all facts:

(25) F: $f_1, f_2, f_3, \ldots$

Consider the following statement:

(26) Q: the list F of all facts fails to have this statement on it.

Suppose Q is false, i.e., Q is on F. This means that Q does not state a fact. So, contrary to our assumption, F is not really an enumeration *of the facts* since it also contains some statement that does not express a fact. Q should thus be removed from F. But if Q is removed from F, then Q becomes true, and it states a fact about F. But this means that F is not a complete list *of all facts*, after all (Rescher, 2006, p. 48). As Rescher clarifies, the "point is that any supposedly complete listing of facts […] will itself exhibit, as a whole, certain features that none of its individual members can encompass. Once those individual entries are fixed and the series is defined, there will be further facts about that series as a whole that its members themselves cannot articulate" (Rescher, 2009, p. 59).

To better see this point, consider now the following statement:

(27) Z: the list F: $f_1, f_2, f_3, \ldots$ is a complete enumeration of all facts.

Z expresses a fact. So, if *F* is a complete enumeration of all facts, *F* should contain the fact expressed by *Z*, i.e., there should be an integer *k* such that:

(28) $Z = f_k$.

Z should thus occupy the *k*-th place on the *F* listing:

(29) f_k = the list *F* takes the form $f_1, f_2, f_3, \ldots f_k, \ldots$

But this would require f_k to be an expanded version of itself, which is absurd. Thus, it is impossible to provide a complete enumeration of all facts, i.e., facts are uncountable (Ibidem, p. 59). If facts are uncountable while the truths that we are able to express are at most countable, there must be some unknowable facts. The same kind of reasoning can be deployed for supporting the claim that metaphysical possibilities are uncountable, while conceivable possibilities are countable, and so that there must be some inconceivable possibilities. Thus, we have some reason to doubt that we are reliable at assessing counterpossibles.

11. Conclusion

In this chapter, I considered the proposal made by Shaffer (2012) and Rice (2021) that scientific realism should be regarded as a commitment to the truth of the modal information provided by scientific models and theories. That proposal aims at reconciling realism with the role played in science by highly idealized models, i.e., those models in which it figures an idealization that introduces such a distortion of the model's target that the model cannot even be interpreted as an approximate true representation of its target. That kind of models seems at odds with the standard understanding of scientific realism as the claim that our best scientific theories and models are true or approximately true. Moreover, an adequate account of scientific practice seems to require the adoption of a non-standard semantics for counterfactuals, i.e., a semantics that admits of impossible worlds. As said, according to Shaffer (2012) and Rice (2021), to rescue scientific realism, one has to understand it as a commitment to the truth of the modal information provided by models. This seems to imply a commitment to some form of modal realism. I focused on just one possible objection that can be made to that variant of scientific realism. To be realist about the modal information provided by models, our assessment of counterpossibles need to be reliable. For our assessment of counterpossibles to be reliable, we have to be able, at least in principle, to examine all possible and impossible worlds. I considered Rescher's arguments for supporting the claim that there are inconceivable possibilities. If there are inconceivable possibilities, there might also be inconceivable impossibilities, and so there might be possible and impossible worlds of which we cannot have knowledge. If this is the case, there would be a principled reason to doubt that our assessment of counterpossibles is reliable. This would suggest that we should refrain to take a realist stance on science on the basis of a commitment to a realist stance on modality. And this would mean that Shaffer's (2012) and Rice's (2021) route to scientific realism might not be viable.

References

Adams, R. (1974). Theories of Actuality. *Noûs*, 8:211–231.

Baker, A. (2009). Mathematical Explanation in Science. *The British Journal for the Philosophy of Science*, 60:611–633.

Baron, S., Colyvan, M., & Ripley, D. (2017). How Mathematics Can Make a Difference. *Philosophers' Imprint*, 17:1–19.

Berto, F., & Jago, M. (2019). *Impossible Worlds*. Oxford: Oxford University Press.

Berto, F., & Jago, M. (2023). Impossible Worlds. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/sum2023/entries/impossible-worlds/>.

Berto, F., French, R., Priest, G., & Ripley, D. (2018). Williamson on Counterpossibles. *Journal of Philosophical Logic*, 47:693–713.

Bueno, O., Menzel, C., & Zalta, E.N. (2014). Worlds and Propositions Set Free. *Erkenntnis*, 79:797–820.

Chakravartty, A. (2017). Scientific Realism. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/>.

Clarke-Doane, J. (2019). Modal Objectivity. *Noûs*, 53:266–295.

Cresswell, M. (1966). The Completeness of S0.5, *Logique et Analyse*, 9:263–266.

Dean, W. (2023). Recursive Functions. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/sum2023/entries/recursive-functions/>.

Dohrn, D. (2023). Modals Model Models: Scientific Modeling and Counterfactual Reasoning. *Synthese*, 201:161.

Dohrn, D. (forthcoming). The Science of Counterpossibles vs. the Counterpossibles of Science. *The British Journal for the Philosophy of Science*, DOI: 10.1086/716769.

Elgin, C. (2007). Understanding and the Facts. *Philosophical Studies*, 132:33–42.

French, R., Girard, P., & Ripley, D. (2022). Classical Counterpossibles. *The Review of Symbolic Logic*, 15:259–275.

Frigg, R., & Hartmann, S. (2023). Models in Science. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/fall2023/entries/models-science/>.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38:173–198.

Goodman, N. (1947). The Problem of Counterfactual Conditionals. *The Journal of Philosophy*, 44:113–118.

Hales, T.C. (2001). The Honeycomb Conjecture. *Discrete and Computational Geometry*, 25:1–22.

Hodges, W. (2013). Modality in Mathematics. *Logique et Analyse*, 56:5–23.

Iranzo-Ribera, N. (2022). Scientific Counterfactuals as Make-Believe. *Synthese*, DOI:10.1007/s11229-022-03949-8.

Jenny, M. (2018). Counterpossibles in Science: The Case of Relative Computability. *Noûs*, 52:530–560.

Kiourti, I. (2010). *Real Impossible Worlds: The Bounds of Possibility*. Ph.D. Dissertation, University of St Andrews.

Kocurek, A.W. (2021). Counterpossibles. *Philosophy Compass*, e12787.

Kripke, S. (1965). Semantical Analysis of Modal Logic II: Non-Normal Modal Propositional Calculi. In J.W. Addison, L. Henkin, & A. Tarski (Eds.), *The Theory of Models* (pp. 206–220). North-Holland: Amsterdam.

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

Kroedel, T. (2012). Counterfactuals and the Epistemology of Modality. *Philosophers' Imprint*, 12:1–14.

Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical? *The British Journal for the Philosophy of Science*, 64:485–511.

Lemmon, E. (1957). New Foundations for the Lewis Modal Systems. *Journal of Symbolic Logic*, 22:176–186.

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

Lewis, D. (1986). *On the Plurality of Worlds*. Oxford: Blackwell.

Linnebo, Ø. (2017). *Philosophy of Mathematics*. Princeton: Princeton University Press.

Mallozzi, A., Vaidya, A., & Wallner, M. (2023). The Epistemology of Modality. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/fall2023/entries/modality-epistemology/>.

Magnani, L., & Bertolotti T. (Eds.) (2017). *Springer Handbook of Model-Based Science*. Cham: Springer.

McLoone, B. (2021). Calculus and Counterpossibles in Science. *Synthese*, 198:12153–12174.

Menzel, C. (2023). Possible Worlds. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/fall2023/entries/possible-worlds/>.

Nolan, D. (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, 38:535–572.

Nolan, D. (2021). Impossibility and Impossible Worlds. In. O. Bueno, & S.A. Shalkowski (Eds.), *The Routledge Handbook of Modality* (pp. 40–48). London-New York: Routledge.

Odenbaugh, J. (2011). True Lies: Realism, Robustness, and Models. *Philosophy of Science*, 78:1177–1188.

Pincock, C. (2012). *Mathematics and Scientific Representation*. Oxford: Oxford University Press.

Post, E.L. (1944). Recursively Enumerable Sets of Positive Integers and their Decision Problems. *Bulletin of the American Mathematical Society*, 50:284–316.

Priest, G. (1997). Sylvan's Box: A Short Story and Ten Morals. *Notre Dame Journal of Formal Logic*, 38:573–581.

Priest, G. (forthcoming). Mission Impossible. In Y. Weiss, & R. Padró (Eds.), *Saul Kripke on Modal Logic*. Springer: New York.

Rescher, N. (2006). *Studies in Cognitive Finitude*. Berlin-Boston: De Gruyter.

Rescher, N. (2009). *Unknowability*. Lanham (MD): Lexington Books.

Rescher, N. (2019). Conceivability. *Humanities Bulletin*, 2:8–19.

Rescher, N. (2020). *Knowledge at the Boundaries*. Cham: Springer.

Reutlinger, A. (2017). Explanation beyond Causation? New Directions in the Philosophy of Scientific Explanation. *Philosophy Compass*, DOI: 10.1111/phc3.12395.

Reutlinger, A., Colyvan, M., & Krzyżanowska, K. (2022). The Prospects for a Monist Theory of Non-causal Explanation in Science and Mathematics. *Erkenntnis*, 87:1773–1793.

Rice, C. (2021). *Leveraging Distortions*. Cambridge (MA): The MIT Press.

Rohwer, Y., & Rice, C. (2016). How are Models and Explanations Related? *Erkenntnis*, 81:1127–1148.

Rowbottom, D.P. (2019). Extending the Argument from Unconceived Alternatives: Observations, Models, Predictions, Explanations, Methods, Instruments, Experiments, and Values. *Synthese*, 196:3947–3959.

Sandgren, A., & Tanaka, K. (2020). Two Kinds of Logical Impossibility. *Noûs*, 54:795–806.
Sankey, H. (2008). *Scientific Realism and the Rationality of Science*. Burlington: Ashgate.
Shaffer, M.J. (2012). *Counterfactuals and Scientific Realism*. New York: Palgrave Macmillan.
Sklar, L. (1981). Do Unborn Hypotheses Have Rights? *Pacific Philosophical Quarterly*, 62:17–29.
Stalnaker, R. (1968). A Theory of Conditionals. In N. Rescher (Ed.), *Studies in Logical Theory* (pp. 98–112). Oxford: Blackwell.
Stanford, K.P. (2006). *Exceeding Our Grasp.* Oxford: Oxford University Press.
Starr, W. (2022). Counterfactuals. In E.N. Zalta, & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/win2022/entries/counterfactuals/>.
Sterpetti, F. (2022). Mathematical Explanations in Evolutionary Biology or Naturalism? A Challenge for the Statisticalist. *Foundations of Science*, 27:1073–1105.
Tan, P. (2019). Counterpossible Non-Vacuity in Scientific Practice. *The Journal of Philosophy*, 116: 2–60.
Tan, P. (2021). Inconsistent Idealizations and Inferentialism about Scientific Representation. *Studies in History and Philosophy of Science*, 89:11–18.
Turing, A.M. (1936). On Computable Numbers, with an Application to the *Entscheidungsproblem*. *Proceedings of the London Mathematical Society*, 42:230–265.
Turing, A.M. (1939). Systems of Logic based on Ordinals. Proceedings of the London Mathematical Society, 2:161–228.
Vander Laan, D. (2004). Counterpossibles and Similarities. In F. Jackson, & G. Priest (Eds.), *Lewisian Themes: The Philosophy of David K. Lewis* (pp. 258–275). Oxford: Clarendon Press.
van Fraassen, B.C. (1989). *Laws and Symmetry*. Oxford: Oxford University Press.
Vickers, P. (2013). *Understanding Inconsistent Science*. Oxford: Oxford University Press.
Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
Williamson, T. (2017a). Counterpossibles in Semantics and Metaphysics. *Argumenta*, 2:195–226.
Williamson, T. (2017b). Semantic Paradoxes and Abductive Methodology. In B. Amour-Garb (Ed.), *Reflections on the Liar* (pp. 325–346). Oxford: Oxford University Press.
Williamson, T. (2018). Counterpossibles. *Topoi*, 37:357–368.
Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research*, 53:1-42.
Yagisawa, T. (1988). Beyond Possible Worlds. *Philosophical Studies*, 53:175–204.
Yagisawa, T. (2010). *Worlds and Individuals, Possible and Otherwise*. Oxford: Oxford University Press.
Yli-Vakkuri J., & Hawthorne, J. (2020). The Necessity of Mathematics. *Noûs*, 54:549–577.
Zalta, E.N. (1997). A Classically-Based Theory of Impossible Worlds. *Notre Dame Journal of Formal Logic*, 38:640–660.