

# Superexplanations for Counterfactual Knowledge

\*\*Final draft—Please cite the published version\*\*

*Philosophical Studies* (2020) <https://doi.org/10.1007/s11098-020-01477-0>

## Abstract

I discuss several problems for Williamson’s counterfactual-theory of modal knowledge and argue that they have a common source, in that the theory neglects to elucidate the proper constraints on modal reasoning. Williamson puts forward an *empirical* hypothesis that rests on the role of counterfactual reasoning for modal knowledge. But he overlooks central questions of *normative* modal epistemology. In order for counterfactual reasoning to yield correct beliefs about modality, it needs to be suitably constrained. I argue that what is needed is, specifically, information concerning the nature or essence of things. By integrating this information, essentialist deduction arguably provides a better account of our knowledge of modality. Furthermore, I argue that essences have distinctive causal and explanatory powers—indeed, essences are *superexplanatory* for how things are. Compared to Williamson’s counterfactual-theory, superexplanatory essentialism clarifies what the proper constraints on modal reasoning are, and why they have such a special status.

## Introduction

Contemporary debates in modal epistemology feature three main families of theories: *conceivability-theory*, *counterfactual-theory*, and *essentialist deduction* (cf. Vaidya 2017; Vaidya and Wallner 2018; Mallozzi 2019). In its broadest lines, *conceivability-theory* holds that our imaginative capacities, when adequately qualified, can yield knowledge of metaphysical possibility and necessity.<sup>1</sup> Within Western thought, elements of conceivability-theory can be found in influential views such as classical Cartesianism—where the source of our modal knowledge is a distinctive kind of *intellectual grasp*—as well as Hume’s cognitive psychology—where our modal beliefs derive from experientially informed imagination. Within recent debates, prominent examples of conceivability-theory include Chalmers 2002; Ichikawa and Jarvis 2012; Kung 2010; Menzies 1998; Yablo 1993 among others (see also the now classic Gendler and Hawthorne 2002). Importantly, conceivability-theorists often endorse some form of *rationalism*, by holding an *a priori* connection between conceivability and possibility. This is controversial. There is a longstanding skepticism against epistemologies that heavily rely on *a priori* methods. One main concern is that such methods are to be carried out by some special, *sui generis* faculty for accessing modal truths (as well as, arguably, mathematical, logical, and normative truths. For discussion, see e.g. Bonjour 1998; Boghossian and Peacocke 2000; Boghossian and Williamson forthcoming; Devitt 2005). Furthermore, the thesis of an *a priori* link between conceivability and possibility is sometimes further conjoined to an *idealized* notion of conceivability (e.g. Chalmers 2002), which has drawn further criticism (e.g. Priest 2016; Roca-Royes 2011a; Worley 2003).

---

<sup>1</sup> I will mostly speak of *knowledge* of modality. However, what I say should be applicable to more modest epistemological targets, such as *justified belief* or *understanding* of modality, and to be neutral enough to accommodate internalist as well as externalist broader epistemological perspectives.

*Counterfactual-theory* constitutes a main alternative to traditional rationalist conceivability-theory. Here I will discuss specifically Williamson's version (2007). (But see also Hill 2006; Kroedel 2012; Kment 2014, 2018; Vetter 2016). Williamson's counterfactual-theory assimilates our capacity for knowing metaphysical possibility and necessity to our capacity for counterfactual reasoning. More specifically, Williamson invokes certain basic logical equivalences involving modal and counterfactual operators, to the effect that statements of possibility and necessity can be reformulated by means of the counterfactual language in a straightforward way. On the basis of those equivalences, Williamson argues that we gain knowledge of what is necessary and possible by evaluating counterfactual suppositions. An advantage of counterfactual-theory compared to rationalist conceivability-theory is that it avoids main criticisms that affect the latter. In the spirit of philosophical *anti-exceptionalism*, Williamson claims that there is no need to appeal to supernatural ideal reasoners, or mysterious *a priori* faculties. Those in his view indicate "a bizarre lack of cognitive economy" (2007: 162) (but see Malmgren 2011 for an argument against this conclusion). For Williamson, modal knowledge can be easily explained within *ordinary* epistemology. Any residual skepticism about metaphysical modal knowledge quickly translates into skepticism about ubiquitous and uncontroversial ordinary counterfactual knowledge. As he puts it, "far from being *sui generis*, the capacity to handle metaphysical modality is an "accidental" byproduct of the cognitive mechanisms that provide our capacity to handle counterfactual conditionals" (162).

On the other hand, conceivability-theory and counterfactual-theory have something important in common, in that they both pursue the issue of how we know about modality by focusing on the investigation of the *means* or *mental operations* that are involved with gaining modal knowledge. In doing so, they adopt a "means-first" approach to modal epistemology, as I call it. Going "means-first" in modal epistemology is in effect to tackle the central question of how we know about modality head-on. For this approach prioritizes individuating and giving an account of the distinctive cognitive/epistemic means and methods involved in modal knowledge, aiming at finding out which faculty or faculties is responsible for our understanding of modality and the formation of modal beliefs. It is thus unsurprising that means-first theories typically feature systematic accounts of our conceivability or imaginative exercises. Indeed, conceivability-theory and counterfactual-theory reach similar answers in this respect (cf. Morato 2017). Williamson himself advocates the epistemic roles of imagination, as he holds that our imaginative capacities are centrally involved in the procedure of developing a counterfactual supposition.

There is a crucial problem that means-first modal epistemologies face. In order for our imaginative procedures to reliably capture modal truth, they need to be somehow adequately *constrained*. Following Vaidya and Wallner (2018), I call this the problem of "Modal Epistemic Friction". There must be some kind of push-back, or *friction*, on modal reasoning to make sure that it does not lead us astray but rather captures genuine possibility and necessity. For our conceivability and other imaginative exercises may present to us a number of (*ceteris paribus*) equally plausible scenarios; while they do not indicate which one is correct. It seems *prima facie* conceivable, for example, that water might have had a different molecular structure, or that Saul Kripke might have been Rudolf Carnap's son. But how do we know from those imaginings alone which ones are *genuinely* metaphysically possible?

Recent literature in the philosophy and psychology of imagination has pointed to a corresponding problem in the case of practical *everyday* imaginative exercises. Even among equally “realistic” imagined scenarios, imagination does not tell you *which one* is the correct one. You can imagine, say, flying rocks as easily as you can imagine rocks falling down because of the actual force of gravity. You can imagine your couch going through the doorway as well as getting stuck in there. All such mental simulations might be equally realistic. Against recent enthusiasms for the epistemic roles of imagination, this can be taken to undermine the thesis that imagination is a source of knowledge (see Spaulding 2016; Langland-Hassan 2016; Kind & Kung 2016).<sup>2</sup>

But what are the relevant constraints for modal reasoning, and in virtue of what do they hold? This may very well be the key problem in trying to explain modal knowledge. I suggest that a profitable way to address it is by adopting what I have called a “*metaphysics-first*” approach to modal epistemology. By contrast with “means-first” theories from the more traditional literature, the metaphysics-first approach stresses that in order to clarify how we know about modality we first need a good grip on what modality is or the nature of modal truth. It thus recommends prioritizing the metaphysical investigation as our metaphysical findings will help us answer the epistemological questions (cf. Mallozzi 2018a, 2019). Particularly, having a grasp on the nature of modality helps us address central normative issues for modal epistemology, such as what the correct modal judgments are and what they depend on. Thus, the metaphysics-first approach indicates a way to handle the problem of Modal Epistemic Friction.

A third leading theory within recent debates notably deploys the metaphysics-first approach to help us cast light on the relevant constraints for modal reasoning. *Essentialist deduction* maintains that those are specifically *essentialist constraints*. While it originally goes back to Kripke’s analysis of the necessary *a posteriori* (1971), essentialist deduction has been developed in systematic form by Lowe 2008, 2012 and Hale 2013, 2018, and has gained increasing popularity in the latest literature. As currently framed, essentialist deduction hinges on specific assumptions in modal metaphysics. According to a widespread Neo-Aristotelian approach to metaphysical modality promoted by Kit Fine 1994, metaphysical necessity depends on facts about essence, so that it is a consequence of the fundamental nature or makeup of things that certain features of things are metaphysically necessary. Knowledge of modality, in turn, proceeds *inferentially*, from premises concerning what is essential to things or part of their nature to conclusions about metaphysical necessity. In other words, the epistemology of metaphysical modality importantly depends on the epistemology of essence. Among recent defenders of essentialist deduction thus understood are, besides myself 2018a, Jago 2018; Tahko 2017; and Vaidya and Wallner 2018.

As I discussed conceivability-theory and how it scores compared to accounts of modal knowledge that appeal to essentialist principles in previous work (2018b) in this paper I focus instead on counterfactual-theory. My aim is to show that essentialist deduction is explanatorily more powerful than counterfactual-theory, and that it avoids a number of problems that arise for counterfactual-theory. As I argue, those problems depend, at bottom, on a failure to elucidate what

---

<sup>2</sup> Note that the criteria for what counts as a *correct* supposition or scenario may thus vary depending on whether we are considering practical everyday possibility, or rather metaphysical possibility, as the relevant constraints may vary (more below). Talk of “correctness” here refers to modal truth broadly understood.

provides the relevant epistemic friction in modal reasoning. Williamson’s method involves keeping certain “constitutive facts” (as he calls them) fixed while developing in imagination the supposition in the antecedent of a counterfactual conditional. He gives us examples of such constitutive facts throughout his discussion, so that we may build a list of candidate constraints on counterfactual reasoning. However, he does not tell us why those facts should be selected rather than others, or what distinguishes constitutive facts from non-constitutive facts. Williamson puts forward an *empirical* hypothesis about the psychological processes and methods that we actually carry out in modal reasoning. But he overlooks the *normative* issues of how or in virtue of what they are constrained, and why. This affects the overall explanatory capacity of counterfactual-theory for modal knowledge. Essentialist deduction, by contrast, provides an account of the relevant constitutive facts, thus addressing the central normative issues. As the relevant constraints are essentialist, modal reasoning is informed and disciplined by facts concerning the nature of things. Moreover, I maintain that we have a principled criterion for individuating those facts among all others. According to the *superexplanatory* account I defend, essences have special explanatory powers for how things are, because they cause many properties and behaviors that typically characterize (instances of) kinds as well as individuals. As such, superexplanatory essentialist deduction is a better option in modal epistemology compared to counterfactual-theory.

### 1. Williamson’s counterfactual-theory

Williamson’s central thesis is that knowledge of metaphysical modality is a “special case” of knowledge of counterfactual conditionals. In particular, “the ordinary cognitive capacity to handle counterfactual conditionals carries with it the cognitive capacity to handle metaphysical modality” (136). The capacity for modal knowledge can be thought in other words as a byproduct of our capacity for counterfactual thinking.

Williamson supports this thesis with an apparatus of logical equivalences between counterfactual operators and modal operators, which show that statements of possibility and necessity can be reformulated in counterfactual terms in a straightforward way. The two central equivalences are the following:

(N)  $\Box A \equiv (\neg A \Box \rightarrow \perp)$   
 (it is necessary that A IFF if it were not the case that A a contradiction would follow)

(P)  $\Diamond A \equiv \neg(A \Box \rightarrow \perp)$   
 (it is possible that A IFF it is not the case that if A were true a contradiction would follow<sup>3</sup>)

---

<sup>3</sup> In quantified contexts, the following two equivalences also hold:

(N)<sub>q</sub>  $\Box A \equiv \forall p (p \Box \rightarrow A)$   
 i.e., something is necessary IFF whatever were the case, it would still be the case, and  
 (P)<sub>q</sub>  $\Diamond A \equiv \exists p \neg (p \Box \rightarrow \neg A)$   
 i.e., something is possible IFF it is not such that it would fail in every eventuality (159).

According to Williamson, (N) and (P) show that “metaphysically modal thinking is logically equivalent to a special case of counterfactual thinking.” Consequently, “modulo the implicit recognition of this equivalence, the epistemology of metaphysically modal thinking is tantamount to a special case of the epistemology of counterfactual thinking” (158). By means of those equivalences, we come to know about possibility and necessity by developing counterfactual suppositions in search for a *contradiction*, while keeping fixed certain *constitutive facts* in the background of our supposition.

Counterfactual reasoning is itself strictly tied to causal thinking, and plays an important, widespread role in everyday life. We resort to counterfactual reasoning in making choices and planning future action, in learning from experience, as well as in interpreting and evaluating other people’s behavior. Also, counterfactual thinking is a key theoretical tool in scientific practice, where hypotheses are often formulated in counterfactual terms: “If it is a law that property P implies property Q, then typically if something were to have P, it would have Q” (141). In sum, “counterfactual thought is deeply integrated into our empirical thought in general” (141).

Importantly, counterfactual thinking is for Williamson largely *imaginative* thinking—and typically, though not necessarily, quasi-perceptual imagining. In assessing a given counterfactual, we evaluate the consequent on the supposition of the antecedent by developing the supposition through an imaginative exercise, which typically involves *offline simulations* of our cognitive capacities. Also importantly, the exercise requires keeping certain *background knowledge* fixed within the scope of the supposition. As a main example of how this works, Williamson considers a case where a rock falls down a mountain and ends up into a bush. What would have happened if the bush had not been there? It is worth quoting Williamson’s answer at length:

A natural way to answer the question is by visualizing the rock sliding without the bush there, then bouncing down the slope into the lake at the bottom. Under suitable background conditions, you thereby come to know this counterfactual:

*If the bush had not been there, the rock would have ended in the lake.*

[...] Somehow, you came to know the counterfactual by using your imagination. That sounds puzzling if one conceives the imagination as unconstrained. You can imagine the rock rising vertically into the air, or looping the loop, or sticking like a limpet to the slope. What constrains imagining it one way rather than another? You do not imagine it those other ways because your imaginative exercise is radically informed and disciplined by your perception of the rock and the slope and your sense of how nature works. The default for the imagination in its primary function may be to proceed as “realistically” as it can, subject to whatever deviations the thinker imposes by brute force: here, the absence of the bush. Thus the imagination can in principle exploit all our background knowledge in evaluating counterfactuals. Of course, how to separate background knowledge from what must be imagined away in imagining the antecedent is Goodman’s old, deep problem of cotenability (1954). For example, why don’t we bring to bear our background knowledge that the rock did not go far, and imagine another obstacle to its fall? Difficult though the problem is, it should not make us lose sight of our considerable knowledge of counterfactuals: our procedures for evaluating them cannot be too wildly misleading. (142-143)

For Williamson, we can generally trust our capacity for assessing counterfactual conditionals to deliver the correct answers, as it is informed and disciplined by a whole lot of pre-existing knowledge concerning how nature works. Our imaginative exercises faithfully trace our cognitive capacities in an “offline” mode; so that they supposedly “inherit”, we may say, the reliability of the ordinary perceptual capacities. Counterfactual reasoning is in other words an important *source* of knowledge, with widespread applications in ordinary life (see Williamson 2016 for the role of imagination in acquiring specifically quotidian, *non-modal* knowledge).

Williamson’s story carries over, via the above-mentioned logical equivalences, a corresponding account of modal knowledge. We gain knowledge of metaphysical modality via the same cognitive capacities and methods at play in the case of counterfactual knowledge. In our example, we know that it is *possible* for the rock to have landed in the lake, since, by imagining away the bush while holding fixed the other details in the perceived scenario and letting our imagination proceed as “realistically” as it can, we won’t reach a contradiction. If Williamson is right, we have a straightforward picture of modal knowledge. In the spirit of philosophical anti-exceptionalism, we no longer need to postulate a special *sui generis* faculty for accessing modal truth. It is worth noting that in Williamson’s view modal knowledge is also not *a priori*, like it has been traditionally characterized. Since counterfactual reasoning typically deploys imagination and offline simulation, the resulting knowledge is rather “hybrid” between traditional *a priori-a posteriori* classifications. As Williamson puts it, modal knowledge is an example of “armchair knowledge”, where experience plays a role that is neither strictly *evidential*, nor merely *enabling* for the resulting knowledge. A strictly evidential role would result in clear cases of *a posteriori* knowledge according to traditional parameters; while a merely enabling role, e.g. acquiring the necessary concepts via empirical means, would not affect the apriority of the resulting knowledge (Williamson 2013. For discussion, see my forthcoming).

## 2. *Descriptive vs. Normative Modal Epistemology*

Here is a problem for Williamson’s account of modal knowledge. Williamson puts forward a speculative *empirical* hypothesis within what we might call *descriptive* modal epistemology—namely, that part of modal epistemology dedicated to individuating belief-formation processes and methods that subjects actually carry out in modal reasoning. But Williamson’s empirical hypothesis does not bear directly on issues of *normative* modal epistemology. This aims instead to elucidate what constrains such processes and methods or in virtue of what those may be correct or incorrect. Normative modal epistemology is importantly concerned with the problem of Modal Epistemic Friction, as mentioned. While Williamson stresses the role of background knowledge of “constitutive facts” for correctly developing the supposition in a counterfactual conditional, it is not clear *what this background information exactly is*, and *why it counts as correct*. As a consequence, his theory of modal knowledge leaves crucial questions of normative modal epistemology unanswered.

I argue that several more specific issues that appear to undermine Williamson’s theory all depend, at bottom, on a failure to address such questions. Let us look at those issues next.

### 3. Issues for Williamson's Counterfactual-theory

#### 3.1 Logical Equivalences and Metaphysical Modality

Williamson's subsumption of the epistemology of modality under the epistemology of counterfactual conditionals relies on logical equivalences (N) and (P). Those are not uncontroversial. In particular, in order for (N) to hold, one has to accept that all counterpossibles (i.e., counterfactual conditionals with impossible antecedents) are vacuously true. But this is actually the subject of heated disputes in the literature, as many have given arguments defending false counterpossibles within semantics that include impossible worlds (e.g. Berto Ripley Priest and French 2017; Brogaard and Salerno 2013; Jago 2013; Nolan 1997; and Restall 1997). Insofar as equivalences (N) and (P) are meant to offer support for Williamson's epistemological reduction, questioning their validity may cast doubt on the broader project.

More generally, it is unclear that logical contradictions directly bear on matters of *metaphysical* possibility and necessity. Counterfactually assuming that water is not H<sub>2</sub>O, or that Saul Kripke is Rudolf Carnap's son, does not entail a logical contradiction. We might accept, perhaps, that there is a formal equivalence between counterfactual statements and modal statements as Williamson draws it. However, such an equivalence does not obviously entail that modal reasoning amounts to a search for contradictions.<sup>4</sup> Moreover, even assuming it does, that might cause the further worry that Williamson's theory may only safely capture *logical* modality, rather than metaphysical modality. For modal reasoning thus understood might seem to only give us a method to establish matters of logical necessity and possibility. (Note that an analogous worry arises for Chalmers' conceivability-theory. However, Chalmers' theory *identifies* logical-conceptual and metaphysical modality, while it also relies on considerations of ideal coherence to access modal truth. See my 2018b).<sup>5</sup>

To sum up this initial worry: while counterfactual-theory gives us a reliable method to grasp possibilities and necessities according to the axioms of logic and basic truth-preserving patterns of inference, it is not clear that the method also secures knowledge of metaphysical modality.

#### 3.2 Constitutive Facts and Pre-Existing Modal Knowledge

One can resist the objection that counterfactual-theory only ranges over logical modality not metaphysical modality. As mentioned, Williamson's method for applying (N) and (P) further involves keeping certain "constitutive facts" fixed within the scope of the supposition while assessing a counterfactual. As Williamson remarks, "the imagination can in principle exploit *all our background knowledge* in evaluating counterfactuals" (2007: 141. My emphasis). Once we integrate background information in our counterfactual supposition, we may effectively go beyond the scope of logical modality and deploy our search for contradictions to establish matters of metaphysical modality. At several points Williamson indicates that knowledge of such constitutive facts should be understood in terms of general knowledge of chemical, physical, and other basic scientific facts, as well as some grasp of the causal and natural laws. More tentatively, background knowledge might

---

<sup>4</sup> Similarly, Jenkins (2008) has argued that, even granted that the logical equivalences hold, it does not follow that we know metaphysical modality through the same cognitive processes by which we know counterfactuals.

<sup>5</sup> Along analogous lines, Ichikawa (2016) points out that Williamson's equivalences could be used to define *any* type of modality.

also include some *essentialist* principles— for example, Kripke’s principle of the *necessity of origin*, which Williamson finds “plausible” (161).

With these extra assumptions in place, developing a counterfactual supposition where, say, water has a different chemical structure than H<sub>2</sub>O, does lead to a contradiction. Likewise it does supposing that Saul Kripke might have been Rudolf Carnap’s son. Whereas, counterfactually supposing that the bush is not there in the falling rock scenario does not.

However, introducing those additional requirements for counterfactual reasoning triggers a number of issues. Several authors, including Roca-Royes (2011a, 2011b) and Tahko (2012), have pointed out that our counterfactual evaluations appear to rely on *pre-existing modal knowledge*. That seems undesirable for the theory, as the method of counterfactual reasoning should instead independently yield modal knowledge. Along similar lines, Boghossian has pointed out that this issue might undermine Williamson’s whole project: “I am very doubtful that knowledge of modal claims can be reduced to knowledge of counterfactuals. It seems to me that, on any plausible account, knowledge of logical, mathematical and constitutive truths will be presupposed in accounting for our knowledge of counterfactuals” (2011: 490, fn. 1).

Williamson might reply in a number of ways. First, he might question the supposed “reduction” of modal knowledge to counterfactual knowledge. He might insist that his thesis is simply that our capacity for handling modal claims is a *special case* or *byproduct* of our capacity for counterfactual thinking. This *per se* does not imply a reduction of modal knowledge to counterfactual knowledge. (It would be of course interesting to push Williamson to clarify further the relationship between the two). Second, Williamson might contend that the counterfactual method need not involve pre-existing knowledge that the constitutive facts are *necessary*. The relevant background knowledge is just general knowledge about how nature works, which we fruitfully put to use in drawing modal conclusions as a product of the counterfactual assessment.

While these considerations might help to an extent, the required background knowledge of constitutive facts leads to even more serious issues.

### 3.3 Knowledge and Analysis of Constitutive Facts

What are the constitutive facts? And how do we select them among all other facts?

As mentioned, we may build a rough list of such constitutive facts based on Williamson’s suggestion. However, Williamson’s theory does not give us any principled *criterion* for individuating those facts. Thus, Tahko (2012) questions how we should select the constitutive facts when that involves deciding between rival scientific hypotheses. For example, how are we to decide whether we should keep fixed atomic number, or rather nuclear charge, in counterfactual reasoning aimed at establishing modal truths involving chemical elements?

Similarly, I find Williamson’s appeal to our “general sense of how nature works” in assessing counterfactual conditionals especially problematic. This might be read as though the causal and natural laws are implicitly built into the supposition, since the laws regulate (or are otherwise roughly responsible for) “how nature works”. Thereby, we would have to keep the laws fixed in our counterfactual developments along with the other background assumptions; in doing so we in effect treat the laws on a par with the “constitutive facts”. While this seems correct when assessing worlds

having the same nomological profile as the actual world, it becomes problematic when we extend our evaluations to more remote metaphysical possibilities. For the natural and causal laws would come out as *necessary*, while the received view seems to be rather that those are metaphysically contingent. One might reply that Williamson's theory certainly allows one to selectively imagine away some particular law of nature while still exploiting a "general sense of how nature works" when developing the supposition<sup>6</sup>. In assessing counterfactuals one shouldn't end up with a contradiction, unless imagining away the law in question is in direct conflict with other details of the supposed scenario. However, it seems a substantive issue whether one could in fact develop the supposition that a certain law might have never been in place, without that potentially impacting other parts of the nomological system and leading to a contradiction. Wouldn't our "general sense of how nature works" be affected by imagining away, say, the actual value of Planck's constant? Or that electrons are negatively charged? The thought here is that even if in principle Williamson's theory should be neutral about necessitarianism, in fact one might need to treat the laws as "constitutive" in order to effectively implement our "general sense of how nature works" in our assessments. On the other hand, Williamson might countenance that the laws are part of a broad "pool" of background information, from which one can draw *selectively* depending on the particular supposition one is considering. Perhaps we might not need to hold fixed *all* the laws for the purpose of evaluating a particular counterfactual. Perhaps certain laws like the ones describing Planck's constant, or the behavior of electrons, are in fact constitutive and couldn't be coherently imagined away, whereas others are not like that. That seems plausible. But then, unless one regards such choices as arbitrary, Williamson owes us some account of the criteria for treating certain laws not others as constitutive, and thus for keeping them fixed in our counterfactual suppositions.

Similarly for other candidate constitutive facts. Take Kripke's essentialist principles, for example. I mentioned that Williamson seems to approve of the necessity of origins; but he is evidently cautious accepting those principles. That makes one wonder what the rationale is for counting certain items as constitutive in specific cases. For example, while Williamson explicitly regards matters such as atomic numbers as constitutive (2007: 164; 170), it is not clear whether that is based on a more general endorsement of the essentiality of fundamental kind, or whether that instead results from somehow individuating "special" facts within chemistry. In sum, the list of constitutive facts that we gather from Williamson's discussion is problematic because it is not clear what it does and does not include, and based on what criteria.

Roca-Royes (2011a, 2011b) pushes an analogous line and argues that for Williamson's method to succeed, we need to know not only the constitutive facts; but also *that they are constitutive*, namely that those are the right facts to be held fixed in our counterfactual supposition. Roca-Royes' objection might be read as a variation on the problem of Modal Epistemic Friction. We need a story of what the correct constraints on modal reasoning are; specifically, we need a story of how we can tell apart constitutive from non-constitutive facts. But there is more to that: it cannot be that we somehow just *happen* to tell those facts apart. Roca-Royes stresses that in order to be able to put those facts to use in counterfactual reasoning, we need to further *know* that those are constitutive facts.

---

<sup>6</sup> Thanks to an anonymous referee for pressing this objection.

From his externalist, reliabilist stand in epistemology, Williamson would likely reply that Roca-Royes sets the bar too high. For one does not need to *know* that those facts are constitutive, and thereby that they count as the right facts to be held fixed in counterfactual reasoning. Along these lines, Yli-Vakkuri has argued for example that it is sufficient that one is somehow reliably *sensitive* to those facts, so that “something short of knowledge is enough” (2013: 619). However, one might insist that relaxing the epistemic requirements in this way only superficially answers the problem. Vaidya and Wallner (2018) point out that the problem re-emerges at the lower level. Even granted more relaxed epistemic requirements, for which, say, the subject only needs to have some kind of epistemic access to some criterion for discerning the constitutive facts, still she needs to be somehow acquainted with those facts in order to create epistemic friction and put them to good use in counterfactual reasoning.

Moreover, one might wonder who are those epistemic subjects doing the counterfactual reasoning that Williamson describes.<sup>7</sup> He seems to be referring to *non-experts*. Apparently, just about any thinker who might find herself engaged with those sorts of reasoning procedures will qualify. Thus, to the extent that people keep certain facts fixed in their ordinary counterfactual suppositions, we (modal theorists) should be able to read off a list of relevant facts out of what they do. But are people reliable concerning these matters? Can we theorists trust the actual practice to track the correct normative constraints? A natural worry is that ordinary subjects’ practice actually misses a lot of things. Also, it seems merely accidental that people think in a certain way rather than another. In sum, there seems to be once again a gap between descriptive empirical modal epistemology and normative modal epistemology that Williamson’s theory needs to address.

Finally, I shall highlight that Peacocke (2011) has questioned the metaphysics underlying Williamson’s counterfactual-theory. Peacocke complains that Williamson does not give us an *analysis* of the constitutive facts and how they are connected to modal facts. At bottom, Williamson does not clarify *metaphysical necessity*. While Peacocke’s criticism is sensible, Williamson might reply that this is not a fair objection. After all, it is not among the aims of his theory to offer an analysis of metaphysical necessity, but only to show that our capacity to handle counterfactual conditionals is sufficient to handle modality as well (cf. Deng 2016). Asking for an analysis of metaphysical necessity goes far beyond the scope of the project. Still, Peacocke’s point helps us stress that for Williamson’s theory to have some serious epistemological bearing as far as normative modal epistemology is concerned, it needs to integrate an account of the underlying metaphysics. While the theory is not required to give us a definitive account of metaphysical necessity, it should at least elucidate the constitutive facts and their connection to modal facts. In general, we can hardly assess claims about how we know about some area of reality without knowing something about the nature of that area. And the more we know about such a nature, the better we can address the epistemological questions. This is indeed the core idea of the “metaphysics-first” approach to modal epistemology (cf. Peacocke 2019; Devitt 2010).

I will argue that essentialist deduction scores better than Williamson’s counterfactual-theory, because it offers a substantive account of the constitutive facts. As we are looking at typical cases of knowledge of metaphysical necessity, the constitutive facts are plausibly *essentialist facts*. Moreover, I

---

<sup>7</sup> Thanks to David Papineau for helpful discussion of this point.

hold that we have a principled criterion for individuating those facts among all others. According to the *superexplanatory* account I defend, essences have special explanatory powers for how things are, because they cause many properties and behaviors that typically characterize individuals and (instances of) kinds. This criterion thus invites us to investigate, case by case, which properties play the relevant causal and explanatory roles for the relevant entities. Before turning to superexplanatory essentialism, however, there is an additional set of issues for counterfactual-theory that deserves attention.

### 3.4 Causal-Nomological Necessity and Metaphysical Necessity

In 3.1 we saw that there appears to be a problem for the purported logical equivalence between modal statements and counterfactual statements, for those equivalences *per se* only give us a method to establish matters of logical not metaphysical modality. Here we are still concerned with whether Williamson's theory adequately handles metaphysical modality; but the objection this time is that his counterfactual method merely leads us to knowing *causal-nomological* modality.

In different ways, Lowe (2012), Deng (2016), Gregory (2017), and Thomasson (2018) all have called into question the capacity of Williamson's account to elucidate knowledge of metaphysical modality. Gregory contends that it is not clear that the same sorts of considerations effectively bear upon mundane counterfactual judgments as well as metaphysical modal judgments. Lowe and Deng similarly argue that Williamson's account only explains knowledge of causal-nomological modality not metaphysical modality. The reason for this is that knowledge of metaphysical modality is subsumed as a special case of knowledge of counterfactual conditionals, where this is in turn strictly causal-nomological knowledge. Lowe worries that Williamson is then "surely misrepresenting the metaphysical modalities as a species of causal modality" (2012: 932). Whereas Deng points to "a gap between the modality involved in our ordinary counterfactual thinking, which is usually causal, and the 'metaphysical' modality properly so-called" (2016: 490). Thomasson (2018) further remarks that this problem not only affects Williamson's theory, but is widespread in contemporary modal epistemology. Other recent empiricist accounts of modal knowledge that are for Thomasson equally open to this objection include Vetter (2015), Bueno and Shalkowski (2015), Leon (2017), and Roca-Royes (2017).

How does Williamson's theory deal with traditional cases of metaphysical possibility and necessity, based on causal counterfactual thinking? How is the theory to answer distinctively metaphysical questions such as, say, whether a zombie-world is possible? (cf. Deng 2016: 480). Van Inwagen (1998) has famously distinguished between *ordinary, everyday* modal knowledge, which we can and often do have, vs. modal knowledge concerning subject matters that are *remote from everyday life*. While the former modal knowledge plausibly comprises everyday causal counterfactual knowledge, the latter is the sort of modal knowledge that is at stake in fanciful philosophical arguments—like the zombie case just mentioned. Van Inwagen is skeptical that this modal knowledge is within our reach. On his part, Williamson explicitly says that his theory is meant to cover the ordinary cases rather than the more extravagant ones. His theory does not wish to speak about such things as philosophical zombies, and how those might be relevant to deciding substantive issues in metaphysics (2007: 164). While ordinary modal knowledge is still by all means

knowledge of *metaphysical* modality, admittedly it only ranges over “close-by” possibilities. But it should be no drawback to the theory that such cases are more tractable than the remote ones. On the contrary, we probably have an additional *practical* interest in explaining the “easy” cases first, as those are mostly directly relevant to the epistemic purposes of our ordinary lives. (Williamson is not alone in restricting the target of modal epistemology in this way. Similarly, Strohming 2015; Leon 2017; and Roca-Royes 2017). On the other hand, one might protest that Williamson’s theory is not especially concerned with the distinctive cases that are of interest in modal metaphysics, whereas those are crucial matters to address for any legitimate theory of modal knowledge. Some would argue that part of the rationale in wanting to pursue the epistemology of modality is precisely to clarify whether and under what conditions we can trust modal reasoning to cast light on those metaphysical matters.

Thus, the issue here is that while Williamson’s theory safely ranges over causal-nomological modality, it is less clear that it can cast light beyond that into the metaphysical realm. Metaphysical possibility seems covered only to the extent that it coincides with causal-nomological possibility; thereby it may remain largely unexplored.

Once again, an analysis of the constitutive facts could help. What sorts of considerations are we to appeal to in developing counterfactual suppositions as Williamson indicates? We know the answer: our “general sense of how nature works”, which is based on a vast amount of background knowledge, and especially on knowledge of chemistry, physics, and other sciences; more generally, we rely on our more or less explicit grasp of the causal laws. This all seems correct. But I stressed in the previous section that it is unclear whether those items are all meant to count as “constitutive facts” in Williamson’s account, with the potential outcome that the actual causal and natural laws come out as necessary. This might not be a draw-back for the theory *per se*. Indeed, it is worth noting that Kripke’s own conception of metaphysical modality, which essentialist deduction largely draws upon, has sometimes been taken to suggest that metaphysical modality might coincide with nomological modality (probably because of certain allusive remarks in *Naming & Necessity*). Still, whether the actual causal and natural laws are metaphysically necessary is not an obvious matter; perhaps, not even one that we may be able to establish via *a priori* philosophical argument. Unless one is overtly eliminativist or deflationist about metaphysical modality, it is an open issue whether causal-nomological modality and metaphysical modality overlap wholly, or in part.

My point here is that candidate theories of modal knowledge that wish to be neutral about the issue of the necessity of laws need the resources to address not only the easy “close-by” cases (where the gap between metaphysical laws vs. causal and natural laws is less obvious), but also the harder “remote” cases, which count as distinctively metaphysical within a non-reductionist framework. It is not clear that Williamson’s theory can do so, especially if the necessitarian interpretation of the nomological requirements is correct. Thomasson has recently made an analogous point, but cashing it out as a general challenge for theories of knowledge of metaphysical modality. “The distinctively *metaphysical* modal features at issue in characteristic metaphysical debates are cases in which we have the *very same empirical information, and same physical laws and properties*, and yet come to different modal conclusions” (2018: 6). Thomasson’s challenge is that appealing to empirical facts and criteria is not sufficient to answer the relevant metaphysical questions. I agree

with Thomasson, and add, in response, that we need to further constrain our counterfactuals evaluations with remote metaphysical content by adding *essentialist* information. The constitutive facts, as mentioned, are arguably essentialist facts, and the sorts of considerations that guide modal reasoning do not just invoke causal-counterfactual principles, but rather *superexplanatory* considerations aimed to capture the nature of things. As I will argue, the latter support distinctively metaphysical Kripkean principles connecting essence and necessity. Against this framework, we have a substantive story of how we can acquire knowledge of *metaphysical* modality. Moreover, independently of whether nomological necessitarianism is true, there is an interesting conceptual difference between causal-nomological and metaphysical modality.

To take stock. The various issues we discussed seem to point to a common source, namely, the failure to elucidate the relevant constraints on modal reasoning. Williamson offers an empirical hypothesis about the cognitive procedures by which we gain modal knowledge; but does not address crucial normative issues including what the constitutive facts are, why they are constitutive, and how we can know them. That limits the explanatory force of his hypothesis as a theory of modal knowledge, while it prompts several objections.

My aim in the remainder of the paper is to show how superexplanatory essentialism is better equipped than counterfactual-theory to address the concerns of normative modal epistemology.

#### 4. *The Superexplanatory Approach to Knowledge of Metaphysical Modality*

Essentialist deduction is based on the idea that we can derive modal knowledge from knowledge of essence. Kripke gave an implicit formulation of this idea in the context of his treatment of *a posteriori* necessities (1971; 1980). He showed how we can come to know that certain empirical and often scientific statements are necessary, on the basis of the conditional: *If P, then necessarily P*. In all of Kripke's examples, "P" stands for some statement that (a) we know to be true via empirical investigation, *a posteriori*; and (b) characteristically involves reference to what is *essential* to a certain individual or kind (for further discussion, see my 2018a, 2018b). In recent work by Hale (2013) and Lowe (2008, 2012), the method of deduction from essentialist truths has been developed in explicit and systematic form. As mentioned, this is in line with Fine's neo-Aristotelian treatment of essence and modality (1994). Fine locates the source of metaphysical necessity in facts about essence. He contrasts this picture with a "modalist" conception that instead analyzes essence in terms of metaphysical necessity, so that the essentialist truths coincide with the metaphysically necessary truths; and necessity is, in turn, a *primitive* that merely captures truth at all possible worlds (analogously for possibility, which captures truth at some possible world). If the Finean conception of metaphysical modality is correct, it is then a suggestion worth exploring that we have a corresponding *epistemic* route available, from knowledge of essence to knowledge of necessity. Of course, it is then a crucial task to clarify *what essences are* and *how we get to know them*.

In my (2018a) I defended the idea that essences are *superexplanatory*. Essences have special *causal* and *explanatory* roles for how things are, which is why they have such a special status within the economy of the properties of a given entity. Specifically, essences cause and explain many properties and behaviors that typically characterize (instances of) kinds, as well as individuals; so that in investigating this special causal structure in the world we may trace it back to those core underlying

properties. For example, the atomic constitution of chemical elements explains why all samples of a certain element have the same density, electrical and thermal conductivity, melting and boiling point, disposition to combine chemically, and so on. More generally, the molecular constitution of any given chemical substance explains why its instances consistently share many properties and behaviors. This superexplanatory structure is thus distinguished from causal-nomological structure more generally by featuring single common causes which give rise to many properties and behaviors, in what we might call “one-to-many” causal networks. It is important to stress the difference between superexplanatory structure and general causal-nomological structure, as that helps address the following worry.<sup>8</sup> One might contend that it is not clear how superexplanatory essentialism scores better than Williamson’s in illuminating knowledge of *metaphysical* modality, given that it is anchored in knowledge of actual-world nomological relations. However, the proposal is not meant to eliminate modal essences in favour of nomological necessity, but rather to explain them in terms of a specific kind of nomological structure. The key suggestion is that superexplanatory one-to-many structure, at least in the range of cases we discussed, explicates the familiar notion of “essence” or the “nature” of things as the source of metaphysical necessity, which features in the Kripkean bridge-principle as well as in Fine’s account as outlined above. If this is correct, whenever we find this one-to-many nomological structure characteristic of a superexplanatory property, that property will be possessed in all metaphysically possible worlds (see also Godman, Mallozzi, and Papineau 2020). In sections 4.1-4.3 below I briefly illustrate a number of applications of superexplanatory essentialism to central cases involving various *natural kinds*, *individuals*, as well as *philosophical zombies*. Here I should stress that the ambition of the account is to gain a better grip on essence or the nature of things by appealing to this specific kind of nomological structure, which in turn helps elucidate knowledge of metaphysical necessity in a wide range of cases.

If this is correct, in most cases we acquire knowledge of essence empirically, largely via scientific investigation aimed at discovering the relevant causal and explanatory information. Then, we may proceed to gain knowledge of metaphysical necessity *inferentially*, on the basis of Kripkean bridge-principles connecting the actual with the necessary. More precisely, these Kripke-conditionals can be treated as instantiations of a basic conditional that connects essence and metaphysical necessity<sup>9</sup>. I call this principle “E”. In the case of kinds,

(E) *If it is essential to x being F that it is G, then necessarily anything that is F is G*

In the case of individuals,

(E)<sub>i</sub> *If x is essentially F, then necessarily x is F*

Finally, at the sentential level, (E) and (E)<sub>i</sub> can be expressed in a straightforward way with the Finean notation:

---

<sup>8</sup> Thanks to an anonymous referee for pushing me on this.

<sup>9</sup> I take principle (E) to be *a priori* like Kripke 1980 originally suggested. For further discussion: Jago 2018; Mallozzi 2018a; and Vaidya and Wallner 2018.

$$(E)_F \quad \Box_x P \rightarrow \Box P$$

which reads, ‘If a proposition  $P$  is true in virtue of the essence of  $x$ ,  $P$  is metaphysically necessary’; where ‘ $x$ ’, depending on the cases, stands for either an individual or a kind.

Importantly, principle (E) holds both at the *metaphysical-constitutive* level and at the *epistemological-normative* level. As to the former: principle (E) expresses the fundamental relationship between essence and metaphysical necessity, thus closing the gap between the actual and the necessary and guaranteeing that if something is essential to a kind or individual, it is also necessary to it. As to the latter: principle (E) guides modal inference to metaphysical necessity, based on this fundamental relationship between essence and metaphysical necessity. Essentialist deduction as cashed out by superexplanatory essentialism seems to give us a straightforward method for modal knowledge.

Superexplanatory essentialism seems to handle well a range of distinctive cases of metaphysical necessity—such as those involving *fundamental kind-membership*, *individual origin*, *the constitution of particulars*, and cases of the *necessity of identity*—and it gives us a robust method to handle new cases as well. We gain knowledge of necessity by inferring appropriately from essentialist truth, that is on the basis of principle (E) and its particular instantiations. And we know about essences empirically, mostly on the basis of scientific investigation. (Note that the account—at least in its present form—does not aim to explain cases of *logical* and *mathematical* necessities, *conceptual* necessities, and *normative* necessities. Further work is required to address such cases where modal knowledge is arguably purely *a priori*).

We thus have an account of the proper constraints for modal reasoning, particularly for those evaluations concerning distinctive cases of metaphysical modality. Compared to Williamson’s counterfactual-theory, the superexplanatory account tells us why we should select certain facts and not others as “constitutive”—namely, because they are *essentialist facts*. Furthermore, the account clarifies *why* they are essential, by pointing to their superexplanatory roles. We have in other words an answer to the problem of Modal Epistemic Friction.

I conclude by giving a few examples of how superexplanatory essentialism, by contrast with counterfactual-theory, clarifies what the “constitutive facts” are and why they count as such, by appealing to the causal and explanatory powers of essences.

#### 4.1 Chemical, Astronomical, and Biological Kinds

*Chemical kinds* offer probably the clearest example of how essences are superexplanatory. Chemical elements as well as compounds, in normal conditions, typically share many, many properties and behaviors. For example, all samples of silver share the same boiling and melting point, the same capacity for electrical and thermal conductivity, the same combinatorial dispositions, and so on. How is that possible? The answer is that all these properties are *caused by* an underlying core property, namely their atomic number (according to a specific subatomic configuration). That is what *explains* all such properties and behaviors. Atomic number thus plays an absolutely unique role within the economy of properties of a sample of silver, which is why it is plausible to single it out as

the essence or “nature” of silver. Accordingly, silver necessarily has atomic number 47. At all possible worlds, atomic number 47 identifies a certain chemical chunk as a sample of silver by actually underlying that superexplanatory one-to-many nomological structure.<sup>10</sup> We can tell a similar story, *mutatis mutandis*, for chemical compounds, such as water, as well as for chemical mixtures, such as rocks and other minerals. Likewise for various kinds of astronomical objects, such as main-sequence stars, white dwarfs, red giants, and so on. Superexplanatory core properties of their internal physical constitution underlie characteristic one-to-many nomological structures, thus qualifying as the essences of such astronomical kinds.

Let us look at *biological kinds*. Traditional essentialism coming from the views of Kripke and Putnam in the 1970s and 1980s is usually taken to claim that biological essences are—not differently from chemical essences—fully *intrinsic*. Kripke argues that tigers have a certain “internal structure”, and that is what in virtue of which, at any possible world, something is a tiger, regardless of its superficial features (1980: 120-121). However, the consensus in the philosophy of biology (insofar as theorists refer to “essences” at all) is to reject that essences are intrinsic. Philosophers of biology think of essences as *relational* and *historical*: what determines kind-membership is having a certain *history*. In Godman, Mallozzi, and Papineau (2020) we take a similar line, arguing that historical and, particularly, *copying* mechanisms have the relevant superexplanatory powers for the many phenotypic features that are typically shared by the members of a given *taxon*; particularly, we look at *species*. While an explanation merely in terms of intrinsic essences fails to deal with non-genetic inheritance and non-sexual reproduction (Papineau and Godman 2020), it is the copying mechanism from common ancestors having certain features that properly explains why all the same features consistently co-occur in the members of a given species. In other words, the essence of a biological species is the *copying mechanism* involving those particular properties because that is what plays the superexplanatory role for all the members of a species.

#### 4.2 Individual Origin

Superexplanatory essentialism might also help us handle metaphysical necessities involving *individuals*.<sup>11</sup> We can treat individuals as similar to historical kinds, by thinking of the different stages of an individual’s history virtually as multiple instances of that individual. Similarly to all the instances of a given kind, all the stages of an individual will share many, many properties. What is that plays the relevant causal and explanatory role for all such properties in this case? By analogy with the biological copying mechanism from common ancestors, it is an appealing hypothesis that it is the *individual’s origin* that plays the relevant roles. Thereby, one’s origin is essential to being a certain individual. Each stage in an individual’s history can thus be thought of as a copy of the previous one, all the way back to the individual’s origin. Although many properties will likely change over time throughout an individual’s history, the individual’s origin will explain all the many stable properties that consistently characterize that individual. As a consequence, necessarily something is a stage of a given individual only if it descends from that particular origin.

---

<sup>10</sup> Of course this does not imply that the many superficial features and behaviors that are caused by the essence are also necessary. Those could easily vary due to environmental conditions (in fact even at the actual world, in exceptional cases, they do), or by assuming that certain laws might have been different.

<sup>11</sup> The ideas in this section are further developed in unpublished work with Marion Godman and David Papineau.

### 4.3 Zombies?

Could there be *zombies*, that is beings that are physically identical to us but that lack consciousness entirely? As is well-known, some think that the answer to this question may settle once and for all the problem of the relationship between the mental and the physical. If zombies are possible, the mental is not fully reducible to the physical. If they are not possible, several options open up regarding the correct account of the relationship between the mental and physical—for example, *identity*, *supervenience*, *grounding*, etc. (the technicalities of this literature won't matter for the present purposes). The zombie case is a chief example of distinctively metaphysical modal speculation. Indeed, it is paradigmatic of conceivability-theory as developed by Chalmers (2002). As mentioned, the zombie case has also been presented as a challenge for counterfactual-theory as developed by Williamson, for we do not seem to be able to settle this case only based on counterfactual thinking (Deng 2016; Thomasson 2018).

How does superexplanatory essentialism handle the zombie case? First of all, I should stress that the account does not claim to have an answer to all the puzzles of modal metaphysics that we might happen to contemplate. By difference with Williamson's counterfactual-theory, however, this does not mean that superexplanatory essentialism sets in advance the range of cases within modal space that it can cover. Particularly, superexplanatory essentialism does not wish to circumscribe its reach to the cases that are closest to the actual world. Superexplanatory essentialism *could* answer many remote and extravagant puzzles of modal metaphysics that we might happen to contemplate, since it gives us a general principled method for answering metaphysical modal questions. As we saw, this involves starting from empirical, typically scientific knowledge of the relevant essentialist superexplanatory facts, and from there inferring to the corresponding necessary truths, based on the basic bridge-principle I called "(E)". Importantly, this account gives us *conditional* knowledge of modality: *if* something is essential, *then* it is necessary. It is empirical, typically scientific work that establishes the truth of the antecedent, by individuating what plays the relevant causal and explanatory roles in the case under consideration. Thus, a preliminary—perhaps disappointing—answer is that superexplanatory essentialism does not claim to have settled the issue of whether zombies are possible—like instead it does with the question whether silver could have had a different atomic number, say.

However, note that the question whether zombies are possible, although obviously modal, is ultimately meant to establish an *actual* truth. Answering this question, as mentioned, supposedly amounts to answering the question whether the mental is nothing "above and beyond" the physical. The speculation over whether *zombies* are possible, if meaningful, may in effect legitimate the method of conceivability-theory for establishing modal truth, en-route to establish an actual truth.

From the point of view of superexplanatory essentialism, wondering whether zombies are possible seems rather an *unnecessary detour* to establish what the right connection between the mental and the physical is. Indeed, the latter issue, although certainly metaphysical, is not a distinctively modal one. We might flip the perspective altogether: it is not *a priori* modal speculation that can give us answers regarding actuality. Rather, it is empirical knowledge that gives us answers regarding what

is possible and what is necessary.<sup>12</sup> Once science will have established what the nature or essence of consciousness is—again, likely by individuating a superexplanatory core for the many features that characterize our conscious activity—we will then be able to infer the relevant modal consequences in the usual way. Whether or not zombies are possible is one of them—though probably a futile one at that point. Like all questions regarding actuality, its answer will mostly depend on empirical results—not on our imaginative exercises.

### *Concluding Remarks*

I argued that Williamson’s counterfactual-theory fails to address normative issues in modal epistemology and, in particular, the problem of Modal Epistemic Friction. Essentialist deduction as cashed out by superexplanatory essentialism, on the other hand, gives us an account of the relevant constraints on modal thinking in terms of essences and their causal and explanatory powers. Moreover, superexplanatory essentialism handles well a number of familiar cases of metaphysical necessity involving both individuals and kinds, thereby offering a stronger account of knowledge of metaphysical modality.

A final suggestion that I wish to outline is that superexplanatory essentialism may not need to be a rival to Williamson’s theory. Instead, it can be thought of as a supplement to it, by offering the kind of story about the constitutive that appears to be missing in Williamson’s theory.<sup>13</sup> Indeed, superexplanatory thinking may play an important role in structuring counterfactual thinking about metaphysical modality. When we suppose away a superexplanatory core of e.g. some natural kind, we are hypothesizing away most of its correlated properties that distinguish that kind from others, as counterfactually supposing away a cause typically requires us to suppose away its effects, too. Counterfactual suppositions could then be developed in virtually any way, and so become theoretically and practically irrelevant. Treating the superexplanatory core fixed as a “constitutive fact” in our evaluations thus secures the correct modal evaluations, like Williamson’s account predicts, while also further clarifying what those facts are and why they have such a special status in counterfactual thinking (see also Godman, Mallozzi, and Papineau 2020).<sup>14</sup>

---

<sup>12</sup> Papineau 2013 makes a similar point. Interestingly, though, he concedes that intuitions of possibility are in general philosophically important not as a source of evidence for or against our theories, but rather as a means to *clarify our thinking*.

<sup>13</sup> Thanks to an anonymous referee for pointing this out.

<sup>14</sup> I am grateful to David Papineau and one anonymous referee at *Philosophical Studies* for helpful comments on a previous draft of this paper. Thanks also to Michael Devitt, Sonia Roca-Royes, Jonathan Schaffer, Anand Vaidya, and Michael Wallner for helpful conversation.

## References

- Berto, F., Ripley, D. Priest, G., and French, R. (2017). Williamson on Counterpossibles. *Journal of Philosophical Logic*, 1-21.
- Boghossian, P. (2011). Williamson on the A Priori and the Analytic. *Philosophy and Phenomenological Research*, 82, 488–497.
- Boghossian, P. and Peacocke, C. (2000). *New Essays on the A Priori*. Oxford: Oxford University Press.
- Boghossian, P. and Williamson, T. (2020). *Debating the a priori*. Oxford University Press.
- BonJour, L. (1998). *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
- Brogaard, B. and Salerno, J. (2013). Remarks on Counterpossibles, *Synthese*, 190, 639–60.
- Bueno, O., and Shalkowski, S. (2015). Modalism and Theoretical Virtues: Toward an Epistemology of Modality. *Philosophical Studies*, 172: 671–689.
- Chalmers, D. (2002). Does Conceivability Entail Possibility? In T. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp.145-201). Oxford: Oxford University Press.
- Deng, D.M. (2016). On the Alleged Knowledge of Metaphysical Modality. *Philosophia*, 44, 479-495.
- Devitt, M. (2005). There is No A Priori. In E. Sosa & M. Steup (Eds.), *Contemporary Debates in Epistemology* (pp. 105-115). Oxford: Blackwell.
- \_\_\_\_\_. (2010). *Putting Metaphysics First*. Oxford: Oxford University Press.
- Fine, K. (1994). Essence and Modality. *Philosophical Perspectives*, 8, 1–16.
- Godman, M. and Papineau, D. (2020). Species Have Historical not Intrinsic Essences. In A. Bianchi (Ed.), *Language and Reality From a Naturalistic Perspective: Themes From Michael Devitt*. Springer Editions.
- Godman, M., Mallozzi, A., and Papineau, D. (2020). Essential Properties are Super-Explanatory. Taming Metaphysical Modality. *The Journal of the American Philosophical Association*.  
<https://doi.org/10.1017/apa.2019.48>
- Gregory, D. (2017). Counterfactual Reasoning and Knowledge of Possibilities. *Philosophical Studies*, 174, 821-835.
- Hale, B. (2013). *Necessary Beings*. Oxford: Oxford University Press.
- \_\_\_\_\_. (2018). Essence and Definition by Abstraction, *Synthese*, 1-17, <https://doi.org/10.1007/s11229-018-1726-7>
- Hill, C. (2006). Modality, Modal Epistemology, and the Metaphysics of Consciousness. In S. Nichols (Ed.), *The Architecture of Imagination: New Essays on Pretense, Possibility and Fiction* (pp. 205–236). Oxford: Oxford University Press.
- Ichikawa, J. (2016). Modals and Modal Epistemology. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*, (pp. 124-144). Oxford: Oxford University Press.
- Ichikawa, J. and Jarvis, B. (2012). Rational Imagination and Modal Knowledge. *Noûs*, 46, 127-158.
- Jago, M. (2013). Impossible Worlds. *Noûs* 47, 713-728.
- \_\_\_\_\_. (2018). Knowing How Things Might Have Been. *Synthese*.
- Jenkins, C. (2008). Modal Knowledge, Counterfactual Knowledge and the Role of Experience. *The Philosophical Quarterly*, 58, 693-701.
- Kind, A. and Kung, P., (Eds.) (2016). *Knowledge Through Imagination*. Oxford: Oxford University Press.

- Kment, B. (2014). *Modality and Explanatory Reasoning*. Oxford: Oxford University Press.
- \_\_\_\_\_. (2018). Essence and Modal Knowledge. *Synthese* 1–23.
- Kripke, S. (1971). Identity and Necessity. In M.K. Munitz, M.K. (Ed.), *Identity and Individuation* (pp. 135-64). New York: New York University Press.
- \_\_\_\_\_. (1980). *Naming and Necessity*. Oxford: Blackwell.
- Kroedel, T. (2012). Counterfactuals and the Epistemology of Modality. *Philosophers Imprint*, 12(12), 1–14.
- Kung, P. (2010): Imagining As a Guide to Possibility. *Philosophy and Phenomenological Research*, 81, 620-663.
- Langland-Hassan, P. (2016). On Choosing What to Imagine. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*, (pp. 61-84). Oxford: Oxford University Press.
- Leon, F. (2017). From Modal Skepticism to Modal Empiricism. In B. Fischer and F. Leon (Eds.), *Modal Epistemology After Rationalism*, (pp. 247-262). Synthese Library. Springer.
- Lowe, E.J. (2008). Two notions of being: entity and essence. *Royal Institute of Philosophy Supplement*, 62, 23–48.
- \_\_\_\_\_. (2012): What is the Source of our Knowledge of Modal Truths? *Mind*, 121, 919-950.
- Mallozzi, A. (forthcoming). Imagination, Apriority, and Modal Knowledge, In *The Epistemic Uses of Imagination*, A. Kind and C. Badura (Eds.), Routledge.
- \_\_\_\_\_. (2019). Introduction, in *New Directions in the Epistemology of Modality*, Special Issue of *Synthese* (2019). <https://doi.org/10.1007/s11229-019-02358-8>
- \_\_\_\_\_. (2018a). Putting Modal Metaphysics First. *Synthese*. <https://doi.org/10.1007/s11229-018-1828-2>
- \_\_\_\_\_. (2018b). Two Notions of Metaphysical Modality. *Synthese*. <https://doi.org/10.1007/s11229-018-1702-2>
- Malmgren, A.S. (2011). Rationalism and the Content of Intuitive Judgment. *Mind*, 120, 263–327.
- Menzies, P. (1998). Possibility and Conceivability: A Response-Dependent Account of Their Connections. In R. Casati (Ed.), *European Review of Philosophy*, 3, 255–277.
- Morato, V. (2017). Conceivability, Counterfactual Thinking and Philosophical Exceptionality of Modal Knowledge. *Topoi*, 1-13.
- Nolan, D. (1997). Impossible Worlds: A Modest Approach. *Notre Dame Journal of Formal Logic*, 38, 535–572.
- Papineau, D. (2013). The Poverty of Conceptual Analysis. In MC Haug (Ed.), *Philosophical Methodology: The Armchair or the Laboratory?* (pp. 166-194). Routledge, Abingdon and New York.
- Peacocke, C. (2011). Understanding, Modality, Logical Operators. *Philosophy and Phenomenological Research*, 82, 472–480.
- \_\_\_\_\_. (2019). *The Primacy of Metaphysics*. Oxford: Oxford University Press.
- Priest, G. (2016). Thinking the Impossible. *Philosophical Studies*, 173, 2649-2662.
- Restall, G. (1997). Ways Things Can't Be. *Notre Dame Journal of Formal Logic*, 38, 583–96.

- Roca-Royes, S. (2011a). Conceivability and De Re Modal Knowledge. *Noûs*, 45, 22–49.
- \_\_\_\_\_. (2011b). Modal Knowledge and Counterfactual Knowledge. *Logique et Analyse*, 54, 537–552.
- \_\_\_\_\_. (2017). Similarity and Possibility: An Epistemology of De Re Possibility for Concrete Entities. In B. Fischer and F. Leon (Eds.), *Modal Epistemology After Rationalism*, (pp. 221-246). Synthese Library. Springer.
- Spaulding, S. (2016). Imagination Through Knowledge. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*, (pp. 207-226). Oxford: Oxford University Press.
- Strohming, M. (2015): Perceptual Knowledge of Nonactual Possibilities. *Philosophical Perspectives*, 29: 363-375.
- Tahko, T. E. (2012). Counterfactuals and Modal Epistemology. *Grazer Philosophische Studien*, 86, 93–115.
- \_\_\_\_\_. (2017): The Epistemology of Essence. In A. Carruth, S. C. Gibb & J. Heil (Eds.), *Ontology, Modality, Mind: Themes from the Metaphysics of E. J. Lowe*. Oxford University Press.
- Thomasson, A. (2018). How Can We Come to Know Metaphysical Modal Truths? *Synthese*.  
<https://doi.org/10.1007/s11229-018-1841-5>
- Vaidya, A. (2017). The Epistemology of Modality. *The Stanford Encyclopedia of Philosophy*, E. Zalta (Ed.), <https://plato.stanford.edu/archives/win2017/entries/modality-epistemology/>
- Vaidya, A. and Wallner, M. (2018). The Epistemology of Modality and The Problem of Modal Epistemic Friction, *Synthese* 1-27.
- van Inwagen, P. (1998). Modal Epistemology. *Philosophical Studies*, 92, 67–84.
- Vetter, B. (2015). *Potentiality: From Dispositions to Modality*. Oxford: Oxford University Press.
- \_\_\_\_\_. (2016). Williamsonian modal epistemology, possibility-based. *Canadian Journal of Philosophy*, 46, 766–795.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell
- \_\_\_\_\_. (2013). How Deep Is the Distinction between A Priori and A Posteriori Knowledge? In A. Casullo and J. Thurow (Eds.), *The A Priori in Philosophy* (pp. 291–313). Oxford: Oxford University Press.
- \_\_\_\_\_. (2016). Knowing by Imagining. In A. Kind & P. Kung (Eds.), *Knowledge Through Imagination*, (pp. 112-23). Oxford: Oxford University Press..
- Worley, S. (2003). Conceivability, Possibility and Physicalism. *Analysis*, 63, 15–23.
- Yablo, S. (1993). Is Conceivability a Guide to Possibility? *Philosophy and Phenomenological Research* 53, 1-42.
- Yli-Vakkuri, J. (2013). Modal Skepticism and Counterfactual Knowledge. *Philosophical Studies*, 162, 605-623.