

# The Mechanics of Thought Experiments

## Contents

<b>1</b>	<b>Modalities in Thought Experiments</b>	<b>2</b>
<b>2</b>	<b>The Radical</b>	<b>2</b>
2.1	Williamson's Positive Account . . . . .	4
2.2	Epistemological Analyticity . . . . .	5
2.3	(POS*) Fails . . . . .	7
<b>3</b>	<b>Fall from Grace</b>	<b>9</b>
3.1	Self-Refutation? . . . . .	10
3.2	Contrapossibles & Non-counterfactuals . . . . .	11
<b>4</b>	<b>Ethics &amp; Logic</b>	<b>14</b>
<b>5</b>	<b>In from the Cold</b>	<b>16</b>
5.1	Paradigm Conceptual Cases . . . . .	16
5.2	Defense Strategies . . . . .	18
5.3	Against (EpAn) . . . . .	18
5.4	Counterfactuals . . . . .	21
5.5	(POS') Works . . . . .	23
<b>6</b>	<b>Summary</b>	<b>25</b>
	<b>References</b>	<b>26</b>

# 1 Modalities in Thought Experiments

The term ‘thought experiment’ (TE) groups together apparently dissimilar analytic processes. One plausible unifying characterisation is that TEs test our intuitions about what could happen. Modality then, is key.

I focus on *philosophical* TEs. These have a rich modal landscape, involving many kinds of modality, e.g. deontic, metaphysical, conceptual. Yet, we seem *not* to have a good pretheoretic handle on the mechanics of successful TEs for any of these modalities. This motivates two questions: how do TEs work (nature), and what are the modal characteristics of TEs (modality)?

The ‘conventional wisdom’ is that TEs in analytic philosophy typically involve conceptual possibility, hence “[make] trouble for one or another professed conceptual analysis” (Jackson, [16]:101). Williamson contests this, arguing that the single primitive kind of modality is metaphysical (modal monism), and that there is no healthy notion of conceptual modality. In §2, I outline his arguments against conceptual modality, along with his uniform account of TEs: that all TEs involve a modal claim and a counterfactual claim.

This account is highly controversial. In §3, I raise some problems faced by it and its justification, targeting the counterfactual analysis and modal monism in turn. These largely stem from Williamson’s over-generalisation from his only case study, Gettier’s TE.

§4 highlights the modal monist’s failure to adequately capture TEs from ethics and logic. But Williamson’s primary target is conceptual modality. Finally, in §5 I defend and develop an alternative account of conceptual modality, and put it to work in a paradigm TE. This lends weight to a plural account of TEs, and perhaps to modal pluralism as well.

## 2 The Radical

Williamson answers the nature and modality questions in one fell swoop. His central hypothesis is that (arguments based on) TEs are best analyzed as consisting of a metaphysical possibility claim and a counterfactual judgement. “[P]aradigm [TEs] in philosophy are simply valid arguments about counterfactual possibilities” ([33]:207). Williamson claims that such counterfactual judgements are a contiguous part of ordinary counterfactual thinking, they are ‘naturalistically innocuous’ ([15]:221).

As motivation, he considers the logical structure of Gettier’s TE [13]. Let

‘ $\diamond_M$ ’ and ‘ $\Box_M$ ’ represents metaphysical possibility and necessity, and take the following dictionary:

$GC(x, p) \equiv$  a Gettier story holds between subject  $x$  and proposition  $p$   
 $JTB(x, p) \equiv x$  has a justified true belief that  $p$   
 $K(x, p) \equiv x$  knows that  $p$

Williamson formalises the modal claim in Gettier’s TE thus:

1.  $\diamond_M \exists x \exists p GC(x, p)$

To reject Williamson’s formalisation of the justified true belief (JTB) analysis of knowledge, that  $\Box_M \forall x \forall p (JTB(x, p) \equiv K(x, p))$ , we need a premise which conditions on Gettier cases. One tempting premise is a strict indicative conditional:

- 2\*.  $\Box_M [\forall x \forall p (GC(x, p) \rightarrow (JTB(x, p) \wedge \neg K(x, p)))]$

The problem here is that “[i]n philosophy, examples can almost never be described in complete detail” ([33]:185). The necessity operator  $\Box_M$  is just too strong. There could be ‘abnormal’ cases of our under-specified Gettier story, i.e. deviant Gettier cases where, somehow, the belief is not justified. This would violate the strict indicative conditional, (2\*). Hence, (2\*) is not the best representation of the TE ([22]:117).

So Williamson must provide a (generalisable) alternative translation of such TEs. Instead of (2\*), he argues that our natural reaction to a Gettier story is to consider what *would* happen *were* it to obtain, i.e. a counterfactual judgement. Taking the counterfactual operator ‘ $\Box \rightarrow$ ’, where  $A \Box \rightarrow B$  means ‘If  $A$  were the case, then  $B$  would obtain’, he argues that the Gettier intuition is better formalised:

2.  $\exists x \exists p GC(x, p) \Box \rightarrow \forall x \forall p (GC(x, p) \rightarrow (JTB(x, p) \wedge \neg K(x, p)))$

He argues that such a counterfactual judgement more accurately describes what occurs in Gettier’s TE. We *judge* how things are, rather than deduce from any linguistic understanding of concepts.

Williamson evidently thinks that TEs feature this ‘judgement of things’ partially *because* our capacities to judge and classify are ‘better’ than our linguistic understanding of concepts ([33]:188), from which we might then deduce classifications. What is meant by ‘better’ is unclear. Linguistic understanding of concepts involves inferential connections, etc. Perhaps he means that this understanding is more complex than being able to judge *how things are*. So, in Gettier’s TE, our *judgement* of whether the particular

belief constitutes knowledge is ‘better’ than our understanding of the concept ‘know’, from which we might deduce whether this belief is known.

Moreover, unlike the strict conditional, the counterfactual (2) excludes the problematic ‘deviant’ counterexamples to (2\*) ([33]:200). In possible world talk, the counterfactual only considers close worlds, whilst ‘ $\Box_M$ ’ considers them all.

Williamson holds that the evaluation of a counterfactual is an act of imagination, constrained by perception, and knowledge of how nature works ([33]:143, [22]:122). He claims that this counterfactual reasoning is a “quite ordinary way of thinking” and “not particularly philosophical” ([33]:180); it is just a species of common sense reasoning [11].

So Williamson jettisons any special faculty of insight into metaphysical possibilities or any philosophy-specific intuition to assess TEs ([33]:177). Canonical TEs involve straightforward counterfactual judgements, the “[s]o-called intuitions involve the very same cognitive capacities that we use elsewhere” ([32]:145). The epistemology and metaphysics of TEs are just the epistemology and metaphysics of counterfactuals, which are in turn ordinary judgements.

Finally, Williamson asserts that this is a *complete analysis* of TEs. So he takes Gettier’s TE to be paradigmatic, and the above morals are generalised across all philosophical TEs: they are fundamentally *metaphysical*. But since there seem to be TEs that use other kinds of possibility, e.g. conceptual or deontic, they must be systematically translated, incorporating them into metaphysical possibility, showing that they do not involve other kinds of modality. We explore this now.

## 2.1 Williamson’s Positive Account

Williamson’s account is radical in three respects. Firstly, he analyzes TEs counterfactually, which conflicts with most prior analyses of TEs ([1], [29]). Secondly, hedging slightly, he insists that TEs ‘principally’ concern only metaphysical modality. The third novelty is his rejection of conceptual modality as healthy or sustainable *tout court*.

According to Williamson, Gettier’s TE shows that some JTB is not knowledge, rather than anything about *our concept* of knowledge ([33]:131). More generally, philosophical TEs are foremost revealing about metaphysics, not our concepts. “[T]he subject matter of philosophy is not conceptual in any distinctive sense. Many epistemologists study knowledge, not just our concept of knowledge” ([33]:211). Even if we learn something about the *concept*

of knowledge, e.g. the concept’s intension,<sup>1</sup> we learn principally about its *nature*.

This puts Williamson at odds with the received wisdom that Gettier’s TE shows “that it is conceptually possible to have [JTB] without knowledge” (Jackson [16]:101), or more generally that TEs yield evidence about the natures of our *concepts*. Williamson subverts this view, claiming that TEs contribute little to our conceptual understanding. In short, TEs are metaphysical.

Williamson argues that *if* there were a sustainable notion conceptual necessity, it would still struggle to adequately capture the TE. One issue is that faced by the strict conditional (2\*). Namely, conceptual possibility allows for deviant cases in which the subject is Gettier-related to the proposition, but the belief isn’t justified: “any reasonable understanding of ... ‘conceptually possible’ [allows for] some abnormal instance of the Gettier case which is not an instance of [JTB]” ([33]:205). So the equivalent formulation, with ‘ $\Box_C$ ’ as ‘conceptually necessity’, also fails:

$$2^{**}. \Box_C [\forall x \forall p (GC(x, p) \rightarrow (JTB(x, p) \wedge \neg K(x, p)))]$$

More generally, Williamson holds that there is simply *no* sustainable notion of ‘conceptually possible’ ([34]:128). He argues that purported explanations of conceptual modality rely on an implausible ‘epistemic conception of analyticity’ ([34]:129). Exactly how conceptual modality is affected is not made clear by Williamson.

The argument for TEs being understood as trading on metaphysical modality is twofold:

- (A) Conceptual modality is somehow unsustainable.
- (B) Even if there were such a notion, it doesn’t effectively capture the information that we gain from such TEs.

I turn to these in turn.

## 2.2 Epistemological Analyticity

Williamson’s metaphilosophical arguments attempt to change the tide on a purportedly unjustified turn towards internalisation and concept-first philosophy. Philosophical methodologists are obsessed with ‘internalising’ philosophy into questions about our concepts ([33]:4). He claims that this is

---

<sup>1</sup>Which Williamson claims only interests ‘concept theorists’, not epistemologists ([33]:206).

inaccurate and untenable. Within this view, his account of TEs results from (A), his rejection of conceptual modality *tout court*. His argument for (A) relies on two premises:

(A1) The epistemological conception of analyticity is incorrect.

(A2) Conceptual modality relies on an epistemological notion of analyticity.

The ‘epistemological conception of analyticity’ is supposed to analyze analyticity (truth solely in virtue of meaning) like so [33]:

**EpAn.** A sentence  $s$  is analytic just in case, necessarily, whoever understands  $s$  is disposed to assent to  $s$ .

For example, “All female foxes are vixens” is regarded as analytic. According to (EpAn), if a speaker doesn’t assent to this sentence under ideal conditions, the failure to assent *constitutes* a failure to understand some constituent term ([33]:73).

This conception of analyticity is his target, though he attacks the wider theory that a disposition to assent to some particular sentence constitutes understanding an expression. So being disposed to assent to “A female fox is a vixen” might constitute understanding ‘vixen’. This complication is suppressed without damage to the sketch of Williamson.

Williamson argues that there are *no* such links between understanding and disposition to assent. He claims that a subject does not need to be disposed to assert any particular sentence in order to understand a given word. Two expressions might share the same meaning, e.g. ‘female fox’ and ‘vixen’, in spite of there being *no* connecting sentences which speakers must be disposed to assent in order to understand the expressions ([33]:128). If this is so, (EpAn) does not adequately analyze analyticity.

Williamson justifies this with two examples of imaginary native English speakers who would reject the vixen-female fox link ([33]:86), due to non-standard accounts of logic and language. The first, Peter, only accepts universally quantified phrases if they are instantiated (giving ‘all’ non-standard semantics), and, being a conspiracy theorist, believes there are no foxes. So, Peter thinks that the sentence is both not analytically true and also *false*. Meanwhile, Stephen has standard semantics for universal quantification, but thinks the predicates are vague, with borderline cases. By subscribing to a three-valued logic, Stephen simply denies that the sentence is true or false. Hence both are not disposed to assent to this sentence, yet it is still analytic. So (EpAn) cannot capture analyticity.

Generalising, Williamson argues that this indicates that there is a general algorithm for generating a counterexample to *any* link between understanding and dispositions to assent to sentences. An expert might reject a candidate link for technical reasons, like a sophisticated but nonstandard theory of logic or language. But such deviations do not demonstrate a lack of understanding.

[Any such link] might be rejected by another expert ... on the basis of a subtle theoretical argument. By this hypothesis [(EpAn)], the expert would be mistaken, but making a subtle theoretical error doesn't constitute linguistic incompetence. (Williamson [33]:94).

No such disposition to assent is required for linguistic understanding. So no such disposition to assent is necessary for the sentence to be analytic. So (EpAn) is not a viable analysis of analyticity.

Next, Williamson seemingly assumes that a conceptual notion of modality requires an epistemological conception of analyticity (A2), though he does not say why. As a generous rational reconstruction, I take it that he means that the only way to understand "conceptual modality" is that some statement is conceptually necessary iff it is analytic in an epistemic sense. This presumably relies on some implicit idea that our conceptual apparatus is somehow linked to our epistemological status: claims appear conceptually possible to someone based on their linguistic understanding of the terms involved. Then, as (EpAn) misdescribes the facts, we cannot sustain a theory of conceptual modality - the house of cards collapses.

### 2.3 (POS\*) Fails

Even if there was a sustainable notion of conceptual modality, Williamson argues that it wouldn't explain the phenomena in TEs. The two problems are:

- (B.I) TEs analyzed via conceptual modality lose much of their interest.
- (B.II) Conceptual possibility "does not interact logically in the required way with counterfactuals that play a central role in [TEs]" ([34]:128).

The idea in (B.I) is that if Gettier's TE is understood as concerning conceptual possibility, the conclusion is only that it is *conceptually* possible to have JTB without knowledge. Ostensibly, this doesn't refute the fact that knowledge *just is* (metaphysically necessarily) JTB:

[Such an argument] does not refute the hypothesis that knowledge is [JTB], of metaphysical necessity, any more than the conceptual possibility of something with atomic number 79 that is not gold refutes the hypothesis that gold just is the element with atomic number 79, of metaphysical necessity. (Williamson [33]:206)

Evidently, Williamson thinks that a conceptual reading of Gettier's TE doesn't say enough about what knowledge *is like*, i.e. its nature. Too little is said about the *instances* of knowledge. Only metaphysical modality can capture such information.

As for (B.II), his analysis of TEs uses the possibility principle, (POS) ([33]:156), that *metaphysical* possibility transmits from antecedent to consequent through counterfactuals. Counterfactuals, however, *don't* transmit conceptual possibility ([33]:205).

Suppose  $(A \Box \rightarrow B)$ , i.e. if  $A$  were true then  $B$  would be true. Suppose then that  $A$  is metaphysically possible (or: there is an  $A$ -world), in such a circumstance  $B$  would obtain, so  $B$  is metaphysically possible ( $A$ -worlds are  $B$ -worlds). Hence, for metaphysical possibility, the following holds:

**POS.**  $(A \Box \rightarrow B) \rightarrow (\Diamond_M A \rightarrow \Diamond_M B)$

Williamson frames the Gettier argument thus: from (1), (2), and (POS), we generate the requisite conclusion:

**3.**  $\Diamond_M \exists x \exists p (JTB(x, p) \wedge \neg K(x, p))$

This contradicts his formalisation of the JTB analysis of knowledge,  $\Box_M \forall x \forall p (JTB(x, p) \equiv K(x, p))$ .

Williamson's criticism is that this argument is simply invalid for conceptual possibility (written ' $\Diamond_C$ '), where we replace  $\Diamond_M$  with  $\Diamond_C$ :

**POS\*.**  $(A \Box \rightarrow B) \rightarrow (\Diamond_C A \rightarrow \Diamond_C B)$

**3c.**  $\Diamond_C \exists x \exists p (JTB(x, p) \wedge \neg K(x, p))$

(POS\*) does not hold in general, so (3c) is an unjustified conclusion. To see (POS\*) fail, take  $H = \text{Hesperus}$  and  $P = \text{Phosphorus}$ . The classic example of conceptual possibility says:

**4.**  $\Diamond_C \neg (H = P)$

**5.**  $\neg \Diamond_C \neg (P = P)$

From the ordinary metaphysical identity ( $H = P$ ) and the logic of counterfactuals ([33]:174), we generate:

$$6. \neg(H = P) \Box \rightarrow \neg(P = P)$$

Suppose for a reductio that (POS\*) holds. From (POS\*) and (6), we deduce:

$$7. \Diamond_C \neg(H = P) \rightarrow \Diamond_C \neg(P = P)$$

From (4) and (7), we deduce:

$$8. \Diamond_C \neg(P = P)$$

Which contradicts (5). So (POS\*) fails.

So (POS\*) cannot power the TE argument in its conceptual form. Hence, Williamson argues that a non-counterfactual premise is required. Allegedly, none is suitable. For example, an indicative conditional of the same shape as (POS) cannot do the necessary work in generating (3c) either ([33]:206). The candidate conditional is:

$$\mathbf{POS^{**}}. (A \rightarrow B) \rightarrow (\Diamond_C A \rightarrow \Diamond_C B)$$

But (POS\*\*) fails, as we have ( $H = P$ ) and  $\neg(H = P) \rightarrow \neg(H = P)$ , which jointly imply:

$$6^*. \neg(H = P) \rightarrow \neg(P = P)$$

Arguing as above then contradicts (4) and (5). So neither candidate can do the work of (POS) for conceptual modality.

### 3 Fall from Grace

Here we deliver the first blow for Williamson's uniform account: that metaphysical possibility is not as healthy as he claims, and that his analysis of TEs is not universally applicable.

### 3.1 Self-Refutation?

Suppose we accept some form of metaphysical modality. We can damage Williamson's metaphysics-only account by showing conceptual modality (assuming a healthy notion of this) to be somehow *prior* to metaphysical modality. In this vein, Jackson argues that conceptual possibility claims are our best reasons for justifying that a claim is metaphysically possible ([16]:106).

If this is so, Williamson's argument is self-refuting: he asserts that a modal claim only involves metaphysical, not conceptual, modality. But to support this requires some further conceptual possibility claim. Hence, Williamson's attack on conceptual possibility itself uses conceptually modal assertions, much in the same way that the relativist is forced to employ absolute statements in denying absolutism.

Jackson justified his claim that Williamson's relies on a background conceptual modality through an allegedly paradigmatic example of an argument for metaphysical necessity:

9. If stuff is gold in the actual world, it is gold in every world in which it appears.

10. If stuff has atomic number 79, it has atomic number 79 in every world in which it appears.

11. Gold has atomic number 79 in the actual world.

So:

12. Gold has atomic number 79 in every world in which it appears.

I.e.:

13.  $\Box_M$  Gold has atomic number 79

The claim is that this is paradigmatic for justifying a metaphysically necessary conclusion. (9) and (10) are metaphysical claims. But, following Kripke, Jackson holds that these are *justified* by our reflecting on the concept 'gold'. Such reflection shows us that change from being gold to not gold would be inconsistent with being the same stuff ([16]:107). This use of conceptual reflection, reasons Jackson, means that *the justifications* of (9) and (10) require conceptual modality.

Meanwhile, premise (11) is an a posteriori discovery about the actual world, and unaffected by non-actual considerations. I.e., it is non-modal (either metaphysically or conceptually).

So, arguments for metaphysical modality rely on (i) something a posteriori, and (ii) a priori claims, themselves justified by conceptual claims. Such paradigm cases can only be informative about metaphysical possibility by way of conceptual possibility. Hence, Williamson requires a background theory of conceptual modality to justify his claims about metaphysical modality.

### 3.2 Contrapossibles & Non-counterfactuals

Williamson’s counterfactuals trade on metaphysical modality, and are judged using ordinary reasoning. Accordingly, TEs aim to refute an alleged necessitated target claim by exhibiting a contradictory metaphysical possibility ([22]:116). Gettier’s TE may fit this pattern, but others do not. Two problem cases are TEs which concern metaphysical impossibilities, and those are not adequately captured counterfactually. These are sketched in turn. We also note that ordinary reasoning may not be strong enough to do the work Williamson asks of it.

The metaphysics-only account of TEs provides no clear way to understand arguments involving genuine contrapossibles, i.e. TEs which involve showing something to be impossible for a proof by contradiction.

First, we expand our data-set to include non-philosophical TEs. Williamson’s metaphilosophical view is that philosophy is an extension of natural science. As such, his purportedly uniform account of TEs should also apply to *scientific* TEs. So we can draw on these TEs to show weaknesses in his account.

The classic scientific reductio is Galileo’s linked weights TE ([24]:341). Suppose that heavier objects fall faster than light objects. Then, by tying a feather to a brick, the system falls more slowly than the brick alone, as the feather drags the brick. But the whole system is heavier than just a brick, so must fall more quickly. So the original assumption is contradictory.

Williamson might attempt a translation using the following dictionary:

$L \equiv$  the laws of physics hold

$H \equiv$  heavier objects fall faster

$F \equiv$  a feather tied to a brick falls faster than the brick

14.  $(L \wedge H) \Box \rightarrow (F \wedge \neg F)$

However, we cannot judge this counterfactual using Williamsonian methods alone, since there are no  $(L \wedge H)$ -worlds. Counterfactual reasoning allegedly works via “imagination, constrained by perception and knowledge of how nature works” (Marconi, [22]:122; [33]:143). But imaginations are typically taken to be limited by metaphysical possibility. Our counterfactual

judgements supposedly judge what would occur in such-and-such a metaphysical world. But if such-and-such a situation is metaphysically impossible, it's not clear how Williamson could judge the counterfactual (as either true or false).

Moreover, we cannot in general contrapose counterfactuals ([19]:36). This prevents us from translating the TE counterfactual argument like so:

$$14^*. \neg(F \wedge \neg F) \Box \rightarrow \neg(L \wedge H)$$

$$15. \Box_M \neg(F \wedge \neg F)$$

So

$$16. \Box_M(L \rightarrow \neg H)$$

Even if this particular instance of contraposition can be justified, this argument is unavailable to Williamson since it deviates from the strict interpretation of his uniform theory of the structure of TEs, in three ways. First, there is an additional step, (14\*). Second, Williamson's story has it that the modal claim concerns the antecedent of the counterfactual, but in the current case, (15) and the antecedent of (14) are unrelated. To assume the expected modal claim ' $\neg\Diamond_M(L \wedge H)$ ' would just be circular.

The final deviation is more serious. It looks like we justify (15) via *logical*, not metaphysical, necessity. Judging ' $(F \wedge \neg F)$ ' does not involve metaphysics. We don't need to consider the possibilities of *reality*. We just use ordinary *logical* reasoning to say that this conclusion is a *logically* impossibility. From this logical impossibility, we then deduce (15), that  $(F \wedge \neg F)$  is metaphysically impossible. This method of justification violates Williamson's metaphysical-only account of TEs.

Williamson counters that the contrapossibles problem is not particular to his account [36]. I am unconvinced: he analyzes TEs as *always* using counterfactual judgements based on informed imagination. A bipartite account which instead explains some TE arguments as instances of logical analysis, as seen in (15), may not suffer from this problem. Such accounts bypass the imagination step, looking directly at the meanings of expressions. This motivates a plural account of TEs, perhaps based on modal pluralism, with separate notions of metaphysical and conceptual modality.

Another issue is that some TEs are better captured non-counterfactually, as they involve *conceptual analysis* rather than a metaphysical possibility of Williamson, or the logical possibility of Galileo's TE. One example is Putnam's Twin Earth, which ostensibly shows that the watery stuff on Twin

Earth isn't really water. We come to this realisation through conceptual analysis of the term 'water'.

It seems that we cannot interpret this TE simply via counterfactual imagination, as Williamson maintains. Replacing water with XYZ would have widespread physical effects. On these grounds, one might doubt that there could be a Twin Earth with the requisite properties ([22]:120). But these impossibilities seem legitimately ignored. Moreover, these impossibilities seem to be *metaphysical*, yet to say that Twin Earth is metaphysically impossible seems to miss the point. The TE has philosophical currency despite this impossibility. Put another way:

[it makes] no use of perception or folk science, we do not execute any simulation, and even imagination appears to play, at most, an auxiliary heuristic role. (Marconi [22]:122).

This conceptual analysis is fundamentally dissimilar from Williamson's imagination-based counterfactual reasoning. Twin Earth can't be interpreted as a case of *mere* imagination. This suggests that the metaphysics-only analysis is too narrow.

A final issue concerns Williamson's account of counterfactual reasoning itself. We might first dispute his claimed continuity between ordinary reasoning and highly theoretical reasoning in counterfactuals. Suppose we ignore the general worry that his epistemology of counterfactuals is too speculative to provide any further support ([4]:491). We might still hold that counterfactual intuitions only works for cases close to our experiences. The faculty of reasoning in cases like "if I had missed, the ball would have hit the window" seem disconnected from our judgement of outlandish philosophical TEs like Twin Earth or Chalmers' P-Zombies. While some TEs are close to such ordinary reasoning (e.g. Gettier's), extreme ones are more remote.

Moreover, the modality actually used in ordinary discourse, typified by 'can', 'must', 'possibly' etc., is apparently weaker than the modality required for Williamson's account. Natural language modality is often epistemic. Even in metaphysical cases, Kratzer argues that natural language modality is relativised to a contextually determined subclass of the full class of possible worlds ([18]:531). Natural language modality, and so ordinary reasoning, *cannot* usually access the full space of possible worlds. This suggests that the metaphysical modality and philosophical counterfactuals are distinctly *unlike* those of ordinary reasoning. Others doubt that any faculty could deal with *all* counterfactuals ([21]:307). Williamson owes us substantial justification as to why this faculty of reasoning would be powerful enough to cope with more abstract cases.

## 4 Ethics & Logic

Even if we accept Williamson’s account for a certain subclass of TEs, we may not expect a uniform account for all philosophical TEs. Plausibly, ‘TE’ might pick out a diverse range of things, even within philosophy.

Yet Williamson hopes to give a uniformly metaphysical account. Generalising from Gettier’s TE, he says that *all* TEs consist of a metaphysical modal claim and a counterfactual claim. This is radical. If he is right, he must give a translation of TEs which apparently use different modalities. Critical examples are deontic, logical, and conceptual modality. The only standard tool to hand is *relativised* modality [27], e.g. translating ‘it is physically possible that  $p$ ’ as ‘it is metaphysically possible that the laws of physics hold and  $p$ ’. But relativisation does not seem sufficient. In the following problem cases, the Williamsonian translations into metaphysical claims are unconvincing.

TEs turning on deontic possibility are a concrete problem for Williamson. There is a rich literature of ethical TEs [7]. Prima facie, these TEs involve an independent deontic modality, which describes some normative fact about how the world *should* be. Deontic necessity is distinctly *unlike* metaphysical necessity, since it is non-alethic [27]. Formalising Singer’s pond TE, we have the deontic claim [26]:

**17.** It is right to save an innocent child.

The Williamsonian response must be to translate these into counterfactual claims. One might try to relativise the counterfactuals either conditioning on the consequent or antecedent ([23]:263):

**17c.** Someone saves an innocent child  $\square \rightarrow$  goodness obtains.

**17a.** The world is ideal  $\square \rightarrow$  someone saves an innocent life.

I do not find (17c) particularly compelling, not least due to the introduction of a new proposition “goodness obtains”: it would be odd to require the reification of goodness (itself controversial) just to translate moral sentences. Meanwhile, for (17a) to be non-circular, we seemingly require a non-moral characterisation of ‘ideal’ ([17]:98). Even if this can be done, sentences like (17a) are false in worlds which are ideal but e.g. there are no innocent lives to save. So they are made false too easily.

Moreover, neither translation is sufficiently directed or local. We want to say that *the action of saving the life* is good. As it stands, (17c) obtains in a close possible world where it *isn’t* right to save a child, but goodness obtains

on some other grounds. So it is made true too easily. Similarly, (17a) doesn't tell us that *the action* is good, just that *the world* that action is part of is good.<sup>2</sup>

Finally, one might introduce a (local) morality operator e.g.  $\mathcal{R}(a) \equiv$  'it is right to perform action  $a$ ', and translate via metaphysical necessitation:

**17p.**  $\Box_M(\mathcal{R}(\text{Someone saves an innocent life}))$ .

But this doesn't seem appropriate either. Firstly, its meaning seems different. We don't consider the whole space of metaphysical possibilities to determine whether an action is moral. These possible worlds are irrelevant in ethical judgement. According to the ethical relativist, what is moral depends on the world. So its metaphysically possible for it not to be right to save the life, hence (17p) is false. For the absolutist, ' $\Box_M\mathcal{R}(-)$ ' doesn't *explain* being moral, it is *implied* by it. Moral claims are either true in all worlds, or false in all worlds. So the box ' $\Box_M$ ' is unnecessary in (17p).

The latter point shows the essential problem with (17p), that it passes the buck. We aim to give an *explanation* of the force of moral modality and ethical TEs, but all of this information is now wrapped up into the as-yet-unanalyzed operator  $\mathcal{R}$ . Translating deontic claims as in (17p) is neither theoretically reductive nor explanatory. We are none the wiser for it.

(17) has a different semantic profile from each candidate translation. So the ethical-deontic propositions are not obviously translatable into relativised metaphysical claims. Instead, a new primitive modality may be required.

Logical modality may also be problematic. Logical necessity is meant to capture the notion in play when e.g. logical consequence is defined as "Necessarily, if the premises are true, the conclusion is true".

However, logical necessity has a different kind of force than metaphysical necessity. Suppose that, by some accident of our ambient class of metaphysically possible worlds,  $\Box_M P$  obtains. This wouldn't mean that  $P$  is logically necessary. That's not how logical necessity works: *it doesn't matter* how the worlds are, we don't need to examine any metaphysical worlds to judge logical necessities. Their necessity is not derived from metaphysics.

Many pre-Williamson authors on TEs take logical modality as primitive ([5]:78). One concrete example are p-zombies, which Wilkes argues (problematically) involve logical possibility ([6] §3.2, [31]:36). Vagueness TEs typically concern concepts, but they have logical conclusions ([9]:315). This suggests that here too the underlying argument trades on logical modality.

---

<sup>2</sup>Radical 'situation' semantics might solve this [10].

Even if we doubt that logical possibility had wide currency in TEs, the mere plausibility of such TEs is problematic for Williamson. If this notion is strictly *stronger* than metaphysical necessity, it cannot be a relativised variety of the former, i.e. there is no metaphysical translation.

Certain TEs do not fit in Williamson’s analysis. If we agree with him that there *are* canonical metaphysical TEs, we might need a plural account of TEs. If a single modal primitive does not suffice to explain these various TEs, this would motivate modal pluralism, with several primitive modalities.

## 5 In from the Cold

Williamson implicitly rejects a range of candidate primitive modalities, but his primary target is conceptual modality. I argue that we can develop a defensible account of this. I first exhibit some paradigm TEs which seemingly use conceptual modality. I then present an argument against (A2), that conceptual modality relies on (EpAn), and sketch a full account.

### 5.1 Paradigm Conceptual Cases

[TEs] about persons or about freedom could hardly be about persons or freedom in themselves, they are first and foremost about our concepts of persons and of freedom.

---

Engel [11]:12

Prima facie, Williamson’s metaphysics-only account seems unnecessarily limitative. We seemingly have canonical TEs which *do* concern conceptual possibility. Take Lewis’ surfeit of deviant perceptual causal chains [20]. Their philosophical foci aren’t the particular possible *instances* of seeing, instead they concern candidate *concepts of seeing*. It doesn’t matter whether one believes they are possible or not, the key is that the *kind of possibility* being discussed is conceptual.

One thought is that analyses of the concepts of some field, e.g. epistemology, are ordinary parts of these fields ([16]:103). Williamson’s response is that so-called ‘concept theorists’ might discover a concept’s intension through TEs ([33]:205). But this response appears to be a get-out clause. It is at odds with the rest of his anti-conceptual position (on TEs and metaphilosophy), and

seems specifically designed as a safety net against such criticism. Instead, I put it that there are paradigmatic examples of TEs essentially based on conceptual possibility, which are central in their disciplines, and not merely of interest to concept theorists.

For personhood TEs, the questions are *not* whether the experiments are metaphysically possible - it simply would not answer the personhood theorist's question to teleport someone, or to split a brain ([28]:198). The key is how we *understand* these cases. It seems reason and understanding here are central, rather than imagination (i.e. counterfactual judgement). The philosophy of personhood splits into two principal questions ([31]:21):

- (i) Identification: whether two beings are one and the same person.
- (ii) Characterisation: what it means to be a person at all.

Identification TEs don't seem to concern metaphysical possibility. The particular instance of a person undergoing a brain split isn't the philosophical force of the TE. Nor is the teleportation TE about whether the event occurs in some metaphysical world. Instead, the philosophical content is "whether to describe the event as one of personal survival or one of death and replacement" (Sorensen, [28]:198), i.e. one about *representation* and the *concepts* in play.

Worse still, characterisation questions are difficult to understand non-conceptually. When testing candidate conditions of personhood, we are principally concerned with whether it is *conceptually* necessary that an entity fulfilling them is a person. Metaphysical coextension (or cointension) of two concepts isn't quite enough. If by some accident of metaphysics, 'personhood' coincided with some proposed analysis in every possible world, e.g. 'psychological continuity', we seem not to want to say that personhood *is* psychological continuity. This would be a fact about modal space, not the concept "personhood". In short, there seems to be something *besides metaphysics* which determines the truth of conceptual analysis.

Nor is personhood a special case. Williamson's claim that the primary content of a TE is 'things in themselves' is untenable. The natural reading of classic philosophical TEs, e.g. Block's look-up table TE is that it shows us is that it is *conceptually* possible to have the responses of an intelligent being without being intelligent ([16]:101, [2]). Even Gettier's TE can be construed as about the concept and categorisation of knowledge, not about the classification of particular instances.<sup>3</sup> In summary, some paradigm philosophical TEs seem to be conceptual.

---

<sup>3</sup>Williamson's defenders sometimes lapse into this view ([15]:233).

## 5.2 Defense Strategies

Next, we directly critique Williamson’s polemic against the conceptual modality analysis of TEs. The easiest claim to combat is (B.I), that TEs lose their interest when interpreted conceptually. The issue is over-generalisation. The interest in Gettier’s TE is plausibly captured by metaphysical possibilities. But for the reasons above, the interest in other TEs is seemingly best captured non-metaphysically.

More difficult are (B.II), that counterfactuals don’t preserve conceptual modality from antecedent to consequent, and (A), the alleged unsustainability of the notion of conceptual modality (due to a ‘reliance’ on (EpAn)). Three possible responses are to give:

Reply 1. A non-counterfactual analysis of TEs.

Reply 2. A variant of (POS) for conceptual possibility, i.e. a kind of conditional which preserves conceptual possibility from their antecedents to their consequents.

Reply 3. An account of conceptual modality which doesn’t rely on (EpAn), the epistemological conception of analyticity.

Reply 1 overhauls Williamson’s account. One might return to earlier accounts of TEs, like those of Bealer, Brown, or Norton. But this seems incompatible with Williamson’s analysis of TEs, as it would involve a serious deviation from his counterfactuals-first account of modality. As this is developed elsewhere [8], it is not our focus.

Reply 2 combats (B.II), and can be justified by extending Yablo’s account of conceptual possibilities (supposing that it is sustainable) and counter *actuals*. In §5.4 and §5.5, I use this to justify a variant principle, (POS’).

Reply 3 targets problem (A). Williamson holds that the main weight of the argument against analysing TEs as using conceptual possibility was the implausibility of conceptual possibility *tout court*, rather than any particular attacks on TEs themselves. In §5.3, I suggest this may be unconvincing.

## 5.3 Against (EpAn)

Williamson’s reasoning for (A), the unsustainability of conceptual modality, relies on the premises (A1) (implausibility of (EpAn)) and (A2) (reliance on (EpAn)).

Williamson’s arguments for (A1), the implausibility of the epistemological conception of analyticity, are long, subtle, and somewhat convincing ([33] chapter 4). But there are several issues here. One is strawmanning. (EpAn) is a very narrow notion of analyticity. He claims that it represents a tradition (of e.g. Dummett, Peacocke, and Boghossian [33]:76), rather the position of a particular author. However, it’s not clear that this alleged tradition is unified or that any of the mentioned authors assent to (EpAn). Instead, it looks like Williamson has specified a *test* for analyticity, and taken it to be an (illegitimate) *conception*.

Williamson arguments for (EpAn) have also been criticised for being unnecessarily technical and chisholming, with increasingly niche arguments concerning variants of the same claim - a proof by volume!

We are promised insight, rigour and courageous precision, but what we get is tens of pages of reflection on the sentences ‘All vixens are vixens’ and ‘Vixens are female foxes’, coupled with the admonition that ‘impatience with the long haul of technical reflection is a form of shallowness’ ([33]:45). (Hacker [14]:340).

Further, one might challenge Williamson’s claim that there is a general algorithm for constructing counterexamples to (EpAn). His original argument *stipulates* that there could be such native speakers of English with suitably deviant linguistic faculties ([33]:85). The general claim is dubious ([4]:495). His examples of speakers with non-classical logics and semantics do not seem generalisable. One example turns on vagueness, which would not apply to e.g. mathematical analytic sentences, the other turns on an implausible semantics ([4]:494). We have only Williamson’s word that an expert-based counterexample can *always* be constructed.

By contrast, Williamson says little about (A2), leaving many ways to respond. (A2) is meant to capture his implicit idea that conceptual modality somehow relies on (EpAn). But this is not clear. I present two issues with (A2). The first issue is non-specificity. He is mute on the connection between (EpAn) and conceptual modality, and woolly on exactly what problem the purported flaws in (EpAn) cause for conceptual modality. The second issue is Marconi’s criticism that Williamson’s argument only applies to a *psychological* notion of being “about a concept”.

Williamson is uncharacteristically indirect on the connection between (EpAn) and conceptual modality. It’s just not clear that this is the *only* way to think of conceptual modality. There are several alternative conceptions

of analyticity, including Carnap’s linguistic conception ([35]:46), and epistemological conceptions which do not obviously rely on assent links ([3]:366, [4]:495). We are owed an argument that conceptual possibility would rely on concrete assent-understanding links in the way he maintains.

Moreover, Williamson doesn’t say exactly what the resultant problem for conceptual modality would be, e.g. is it vague, or unintelligible, etc.? Despite claiming that “there is no such thing as conceptual necessity” ([34]:128), he gives no diagnosis of putative accounts of conceptual modality, nor a translation of what happens in putative cases of conceptual possibility. It is a diagnosis without a therapy.

Second is Marconi’s criticism of Williamson, that Williamson’s attack only affects *psychological* accounts of concept. Yet some cases of being ‘about concepts’ have non-psychological analyses (perhaps including mathematical concepts). These elicit an undamaged non-psychological notion of conceptual modality ([22]:123). Marconi then claims that such non-psychological interpretations are difficult to differentiate from being ‘about things’.

This division of concepts can be defended by thinking about how concepts and conceivability interact. Williamson takes conceivability to be wholly psychological ([33]:135). But some concepts are not obviously conceivable, meaning that e.g. Descartes’ 1000-gon would be conceptually impossible. This is clearly wrong. Hence, either conceptual possibility does not rely on conceivability, or conceivability is not entirely psychological.

Philosophers like Ryle and Wittgenstein hold the second line, taking conceivability to be something like *logical possibility* ([22]:113). This elicits a non-psychological notion of concept, and in turn a non-psychological notion of conceptually possible.

More generally, we can make the distinction between psychological and non-psychological concepts more concrete. Take the study of certain psychological maladies, like those ones which inhibit people’s understanding of their own bodies. When we inquire about “pathologies which tend to deprive us of the concepts of body parts ...[, we take] concepts to be categorising functions of the human mind” (Marconi [22]:113). These are psychological concepts. Such concepts are part of psychology proper, and are clearly constituted by the mind.

But other concepts are different. One might consider whether *that thing is a chair*. This consideration is indistinguishable from considering whether *that thing falls under the concept ‘chair’*. These seem non-psychological: the mind has no constitutive role in the second consideration, nor does it in the

first. Thinking of concepts in this way suggests that a TE could be ‘about things’ and ‘about concepts’ simultaneously ([22]:114).

Bringing these together, we conclude that, for non-psychological concepts, conceptual modality need not rely on (EpAn). Indeed, Williamson’s ‘about things’ motivation for metaphysical analysis of TEs also motivates a conceptual analysis in such cases.

## 5.4 Counterfactuals

Given our healthy nascent notion of conceptual modality, the next step is developing a full account to construct some useful machinery for analysing TEs. To do this, we extend Yablo’s theory of counterfactuals [37]. Conceptual and metaphysical possibility are classically differentiated thus [12]:

- It is metaphysically, but not conceptually, possible that the metre stick exists without being a metre long.
- It is not metaphysically, but is conceptually, possible that Hesperus exists without Phosphorus existing.

Yablo’s idea is that for a conceptual possibility, we consider the hypothesis *as though it were actual* even if we do not believe the hypothesis obtains. He calls this a *counterfactual*. “Evaluating S with respect to a counterfactual is asking whether S holds on the hypothesis that *w* is ... this very world.” ([37]:449). He captures this notion with the locution “had it *turned out* that *P*, it would have turned out that *Q*”.

This theory predicts the results of the illustrative examples. Had it turned out the metre stick existed, it would have turned out to be a metre long. So this example is not conceptually possible. Meanwhile, had it turned out that Phosphorus didn’t exist (because there turned out to be no Venus) and e.g. the Hesperus-appearances turned out to be caused by Mars, it would have turned out that Hesperus existed without Phosphorus. Again, the ‘turned out’ locution captures the divergence from Williamson’s counterfactual. So, the nascent notion of conceptual possibility is suitably described by the turned-out locution.

Yablo, contra-Kripke, argues that the turned-out locution is *not* a mistaken case of a counterfactual, as the candidate counterfactual translations are falsified by different contextual settings ([37]:452). In which case, it seems these two notions of possibility do not reduce to one. So we need an independent explanation of counterfactuality.

Next, the counterfactual ‘turned out’ locution must be explained. We can do this non-reductively, with a counterfactual primitive: Williamson helped himself to a primitive ‘ $\Box \rightarrow$ ’, so could we. Let ‘ $P \blacksquare \rightarrow Q$ ’ capture Yablo’s locution ‘had it turned out that  $P$ , it would have turned out that  $Q$ ’.

Next, we need semantics for the conceptual possibility operator. One method is to give Lewisian semantics ([37]:454). Instead, we copy Williamson’s schema for counterfactuals and metaphysical modality. He defines metaphysical possibility like so  $\Diamond_M \phi \equiv \neg(\phi \Box \rightarrow \neg\phi)$  [30], so we could define conceptual possibility like so  $\Diamond_C \phi \equiv \neg(\phi \blacksquare \rightarrow \neg\phi)$ .

Finally, we explain counterfactual judgement. Williamson tied counterfactual judgement to an ordinary faculty of reasoning. Yablo could seemingly do the same. One possibility would be to say that *the very same* kind of common sense reasoning that Williamson latches onto for judging counterfactuals also covers counterfactuals. Williamson provides no argument to think that the ordinary reasoning which he relies on corresponds to the judgement of exactly one kind of modal situation. It’s plausible that the ‘ordinary reasoning’ he speaks of could cover judgements of both counterfactuals and counterfactuals, especially given that typical counterfactuals seem no more obscure and philosophical than typical counterfactuals.

Such non-reductive analyses result in modal pluralism, with independent notions of “conceptual possibility” and “metaphysically possible”, which are two different notions requiring independent explanations. But this is unsurprising: it is exactly what is expected if conceptual TEs cannot be analyzed as metaphysical.

We tread with caution. Yablo conceives of conceptual necessity as *a type* of analyticity ([37]:456). We wish to avoid reliance on (EpAn). In Yablo’s theory, a sentence is conceptually necessary when it is true no matter how things had turned out. This has no prima facie reliance on the treacherous understanding-assertion links to which Williamson objects. So there is no immediate reliance on (EpAn).

Yablo’s account outlines a plausible framework for conceptual modality via counterfactuals. The typical diagnosis of a large proportion of philosophical TEs is that they involve conceptual possibility (e.g. those of personhood or freedom). Naturalism and explanatory efficiency support the analysis of these TEs as featuring the kind of modality they appear to feature, rather than unintuitive metaphysical translation. This lends credibility to the conceptual modality analysis of TEs *so long as* we can give a suitable translation for this class of TEs. This is the task of §5.5.

## 5.5 (POS') Works

Williamson's analysis of Gettier-type TEs relies on (POS), which says that metaphysical possibility is preserved from a counterfactual's antecedent to its consequent. This suggests a schema: perhaps there is a similar premise for conceptual possibility. Replacing metaphysical possibility with conceptual possibility, (POS\*), failed. But there is now a new candidate, which replaces the counterfactual with a Yablo-esque counterfactual:

$$\mathbf{POS'}. (A \blacksquare \rightarrow B) \rightarrow (\diamond_C A \rightarrow \diamond_C B)$$

(POS') avoids the problems Williamson claims affect any conceptual formalisation of TEs. The problem with (POS\*) was that (6) held. Consider the equivalent expression, replacing ' $\square \rightarrow$ ' with ' $\blacksquare \rightarrow$ ':

$$\mathbf{6^{**}}. \neg(H = P) \blacksquare \rightarrow \neg(P = P) .$$

If (6\*\*) held, we would derive the same contradiction as in §2.3. But it doesn't. In fact, had it turned out that Hesperus was not Phosphorus (if, e.g. Phosphorus-appearances were of Mars, not Venus), this would tell us nothing about Phosphorus' non-identity. So (6\*\*) is false.

Moreover, some instances of (POS') are true. It's conceptually possible that Hesperus isn't Phosphorus. Had it turned out that they were different, at least one of them would have turned out not to be Venus. This justified three premises, with  $V \equiv$  Venus:

$$\mathbf{18.} \diamond_C \neg(H = P)$$

$$\mathbf{19.} \neg(H = P) \blacksquare \rightarrow (\neg(H = V) \vee \neg(P = V))$$

$$\mathbf{20.} \diamond_C (\neg(H = V) \vee \neg(P = V))$$

Putting the pieces together, the relevant instance of (POS') is true.

$$\mathbf{21.} [\neg(H=P) \blacksquare \rightarrow (\neg(H=V) \vee \neg(P=V))] \rightarrow [\diamond_C \neg(H=P) \rightarrow \diamond_C (\neg(H=V) \vee \neg(P=V))]$$

However, is (POS') generally true?

Here's a first argument that it is. Suppose that  $A \blacksquare \rightarrow B$ , i.e. had it turned out that  $A$  it would have turned out that  $B$ . So, had it *actually* been that  $A$ , it would have actually been that  $B$ . Now suppose that it is conceptually possible that  $A$ , i.e. it could have turned out that  $A$ . But in which case, it could have turned out that  $B$ . So, the fact that it *actually* could be  $B$  (on the condition that  $A$ ) means that  $\diamond_C A$  ensures  $\diamond_C B$ , i.e. (POS') holds.

Williamson might respond that these intuitions are weaker than those used when theorising about counterfactuals. One explanation of this is that we are less familiar with theorising about counterfactuals than we are with theorising about counterfactuals. But this is just a sociological fact, about the amount of work philosophers have put into analysing counterfactuals.

But let's take this criticism seriously, that the general case strains our modal intuitions. So we might be unable to justify (POS') through intuition. However, any such limitation apparently applies equally well to Williamson's justification for (POS). To demonstrate this, the following is Williamson's argument for (POS), followed by 'mutatis mutandis' argument for (POS'):

Suppose that if  $A$  had held, then  $B$  would have held; then if  $A$  could have held,  $B$  could have held. In terms of worlds: if any closest  $A$  worlds are  $B$  worlds and there are  $A$  worlds, then there are  $B$  worlds. (Williamson, [33]:156)

Hence (POS) holds.

Suppose that if  $A$  had turned out to be the case, then  $B$  would have turned out to be the case; then if  $A$  could have turned out to be the case,  $B$  could have turned out to be the case. In terms of conceptual worlds: if any world that turned out to be an  $A$  world would have turned out to be a  $B$  world and there are conceptually possible  $A$  worlds, then there are conceptually possible  $B$  worlds.

Hence (POS') holds.

Both arguments have the same form, and no extra argument is provided in either case. Further, I put it that the notions involved in each case are only as intuitive as in the other. So the argument for (POS') seems exactly as strong as the argument for (POS). If we accept Williamson's argument for (POS), we have good reason to believe in (POS') also.

Given (POS'), we can analyze conceptual TEs just as Williamson analyzes metaphysical TEs. For example, one interpretation of the teleportation TE is that it shows physical and psychological similarity is insufficient for being described as the same person ([25]:200). We can now capture this TE, using this dictionary:

$P(x, y) \equiv x$  is the same person as  $y$   
 $T(x, y) \equiv x$  enters the teleporter and  $y$  leaves it  
 $A(x, y) \equiv x$  and  $y$  are exactly physically and psychologically alike

22.  $\Box_C \forall x \forall y (P(x, y) \leftrightarrow A(x, y))$

(22) is the target analysis of the concept *person*, which the TE invalidates. This candidate analysis would *not* be appropriately captured using metaphysical necessity for the reasons explained above, principally that metaphysical necessity claim says too little about the *concept*, e.g. it could be validated by an accident of the space of metaphysical worlds.

23.  $\exists x \exists y T(x, y) \blacksquare \rightarrow \forall x \forall y (T(x, y) \rightarrow (A(x, y) \wedge \neg P(x, y)))$

24.  $\Diamond_C \exists x \exists y T(x, y)$

(24) captures the intuition that teleportation is conceptually possible. (23) is the candidate counterfactual judgement that had there turned out to be a teleportation case, it would have turned out that any such case would be such that  $x$  and  $y$  had exactly alike, both physically and psychologically, but are different people. Then by (POS') on (23) and (24), we generate (à la [33]:187):

25.  $\Diamond_C \exists x \exists y (A(x, y) \wedge \neg P(x, y))$

This contradicts (22). This seems to accurately describe the content of the teleportation TE, in denying an alleged analysis of the concept *person*, (22), which metaphysical modality cannot capture.

Significantly, *it doesn't matter* whether one thinks the argument is sound. *Even if* someone thinks that this TE doesn't disprove the candidate analysis of personhood, they accept that it concerns conceptual modality. The rôle of the TE is to provide evidence for our meta-analysis of TEs as sometimes using conceptual modality, *irrespective* of whether it actually disproves a putative analysis. Even in denying (23), one does not deny that the TE concerns conceptual modality, and has a conceptual conclusion.

## 6 Summary

Williamson's uniform 'modal claim-counterfactual claim' account of TEs is inextricably linked to his account of modality, which takes metaphysical modality to be primitive. But he suffers from over-generalising from Gettier's TE. There are plausible candidates TEs which use deontic and conceptual possibilities. Moreover, there is a plausible systematic account of conceptual modality, which is schematically similar to Williamson's account of metaphysical modality. This adequately captures paradigm conceptual TEs where Williamson's account falters. A rich avenue of further investigation would be into other candidate modalities, e.g. whether ethical TEs could be analyzed using a (primitive) deontic logic via Williamson's schema.

## References

- [1] BEALER, G. Intuition and the autonomy of philosophy. In *Rethinking Intuition*, DePaul and Ramsey, Eds. Rowman and Littlefield Publishers, Maryland, 1998.
- [2] BLOCK, N. Psychologism and behaviorism. *Philosophical Review* 90 (1981), 5–43.
- [3] BOGHOSSIAN, P. A. Analyticity reconsidered. *Nous* 30 (1996), 360–391.
- [4] BOGHOSSIAN, P. A. Williamson on the “a priori” and the analytic. *Philosophy and Phenomenological Research* 82 (2011), 488–497.
- [5] BROOKS, D. H. The method of thought experiment. *Metaphilosophy* 25 (1994), 71–83.
- [6] BROWN, J. R., AND FEHIGE, Y. Thought experiment. In *SEP*, Zalta, Ed. 2014.
- [7] BRUN, G. Thought experiments in ethics. In *The Routledge Companion to Thought Experiments*, Stuart, Fehige, and Brown, Eds. Routledge, 2017, pp. 195–210.
- [8] DOHRN, D. Thought experiments without possible worlds. *Philosophical Studies* 175 (2018), 363–384.
- [9] DUMMETT, M. Wang’s paradox. *Synthese* 30 (1975), 201–32.
- [10] ELBOURNE, P. D. *Situations and individuals*. MIT press, 2005.
- [11] ENGEL, P. Philosophical thought experiments: in or out of the armchair? In *Thought Experiments in Historical and Methodological Contexts*, Ierodiakou and Roux, Eds. Brill, 2011.
- [12] GENDLER, T. S., AND HAWTHORNE, J. Introduction. In *Conceivability and Possibility*, Gendler and Hawthorne, Eds. OUP, 2002, pp. 1–70.
- [13] GETTIER, E. Is justified true belief knowledge? *Analysis* 23 (1963), 121–123.
- [14] HACKER, P. A philosopher of philosophy. *The Philosophical Quarterly* 59 (2009), 337–348.

- [15] ICHIKAWA, J., AND JARVIS, B. Thought-experiment intuitions and truth in fiction. *Philosophical Studies* 142 (2009), 221–246.
- [16] JACKSON, F. Thought experiments and possibilities. *Analysis* 69 (2009), 100–109.
- [17] JACKSON, F. *Language, Names, and Information*. Wiley, 2010.
- [18] KRATZER, A. What ‘must’ and ‘can’ must and can mean. *Linguistics and Philosophy* 1 (1977), 337–355.
- [19] LEWIS, D. K. *Counterfactuals*. Blackwell, 1973.
- [20] LEWIS, D. K. Veridical hallucination and prosthetic vision. *Australasian Journal of Philosophy* 58 (1980), 239–249.
- [21] MALMGREN, A. Rationalism and the content of intuitive judgements. *Mind* 120 (2011), 263–327.
- [22] MARCONI, D. Philosophical thought experiments: The case for engel. *Philosophia Scientiæ. Travaux d’Histoire Et de Philosophie des Sciences* 21 (2017), 111–124.
- [23] MONTAGUE, R. Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry* 3 (1960), 259–269.
- [24] NORTON, J. D. Are thought experiments just what you thought? *Canadian Journal of Philosophy* 26 (1996), 333–366.
- [25] PARFIT, D. *Reasons and Persons*. OUP, 1984.
- [26] SINGER, P. Famine, affluence, and morality. *Philosophy & public affairs* (1972), 229–243.
- [27] SMILEY, T. Relative necessity. *The Journal of Symbolic Logic* 28 (1963), 113–134.
- [28] SORENSEN, R. A. *Thought Experiments*. OUP, 1992.
- [29] SOSA, E. Minimal intuition. In *Rethinking Intuition*, DePaul and Ramsey, Eds. Rowman and Littlefield Publishers, Maryland, 1998.
- [30] STALNAKER, R. A theory of conditionals. *American Philosophical Quarterly* (1968), 98–112.

- [31] WILKES, K. V. *Real People: Personal Identity without Thought Experiments*. OUP, 1988.
- [32] WILLIAMSON, T. Philosophical ‘intuitions’ and scepticisms about judgement. *Dialectica* 58 (2004), 109–153.
- [33] WILLIAMSON, T. *Philosophy of Philosophy*. OUP, 2007.
- [34] WILLIAMSON, T. Replies to Kornblith, Jackson and Moore. *Analysis* 69 (2009), 125–135.
- [35] WILLIAMSON, T. *Modal Logic as Metaphysics*. OUP, 2013.
- [36] WILLIAMSON, T. *Counterfactuals and Thought Experiments*, 2016.
- [37] YABLO, S. Coulda, woulda, shoulda. In *Conceivability and Possibility*, Gendler and Hawthorne, Eds. OUP, 2002, pp. 441–492.