Philos Stud (2021) 178:713–729
https://doi.org/10.1007/s11098-020-01453-8

Sensitivity, safety, and impossible worlds

Guido Melchior[1]

Published online: 21 April 2020

Abstract Modal knowledge accounts that are based on standards possible-worlds semantics face well-known problems when it comes to knowledge of necessities. Beliefs in necessities are trivially sensitive and safe and, therefore, trivially constitute knowledge according to these accounts. In this paper, I will first argue that existing solutions to this necessity problem, which accept standard possible-worlds semantics, are unsatisfactory. In order to solve the necessity problem, I will utilize an unorthodox account of counterfactuals, as proposed by Nolan (Notre Dame J Formal Logic 38:535–572, 1997), on which we also consider impossible worlds. Nolan's account for counterpossibles delivers the intuitively correct result for sensitivity i.e. S's belief is sensitive in intuitive cases of knowledge of necessities and insensitive in intuitive cases of knowledge failure. However, we acquire the same plausible result for safety only if we reject his strangeness of impossibility condition and accept the modal closeness of impossible worlds. In this case, the necessity problem can be analogously solved for sensitivity and safety. For some, such non-moderate accounts might come at too high a cost. In this respect, sensitivity is better off than safety when it comes to knowing necessities.

Keywords Sensitivity · Modal epistemology · Safety · Counterfactuals · Impossible worlds

✉ Guido Melchior
guido.melchior@uni-graz.at

[1] Department of Philosophy, University of Graz, Heinrichstrasse 26/5, 8010 Graz, Austria

1 Modal knowledge accounts

Modal knowledge accounts are externalist in nature. They accept that a subject S knows that *p* if her belief that *p* is properly connected to the truthmaking fact and that this connection can be cashed out in terms of counterfactuals. Nozick (1981) argues that S knows that *p* iff S's true belief that *p* tracks truth. Nozick also argues that a modal theory of knowledge is flawed if it does not take the belief forming method into account. Nozick (1981, p. 179) defines knowing via a method as follows:

S knows, via method (or way of believing) M, that *p* iff

(1) *p* is true
(2) S believes, via method or way of coming to believe M, that *p*
(3) If *p* were false and S were to use M to arrive at a belief whether (or not) *p*, then S wouldn't believe, via M, that *p*
(4) If *p* were true and S were to use M to arrive at a belief whether (or not) *p*, then S would believe, via M, that *p*.[1]

Nozick is not particularly clear about his terminology. In line with orthodox terminology, I will call condition (3) the *sensitivity condition* and condition (4) the *adherence condition*.

Nozick's knowledge account is confronted with well-known objections. First, there are instances of insensitive knowledge, as Vogel (1987) and Sosa (1999) point out. Thus, sensitivity is plausibly not necessary for knowledge in contrast to what Nozick claims. Second, sensitivity accounts lead to highly implausible instances of closure failure, as Kripke (2011) shows. These are instances of closure failure that even sensitivity theorists who accept closure failure in the skeptical case reasonably have to reject.[2] As a reaction to these problems, Sosa suggests replacing the modal concept of sensitivity by safety. Sosa's original definition of safety does not take the belief forming method into account. Here is an adapted version of method-relative safety:

If S were to believe that *p* via method M, then *p* would be true.

These are the three modal conditions on knowledge discussed in the literature—sensitivity, adherence, and safety.[3] Orthodox semantics for counterfactuals,

[1] Nozick argues that given this definition of knowing via a method, S knows that *p* simpliciter iff there is one dominant belief forming method, a method that outweighs the other method, and that fulfills conditions (3) and (4). These subtleties will not concern us here.

[2] For a defense of Nozick's tracking theory against Kripke's objection, see Adams and Clarke (2005).

[3] Sensitivity differs from adherence and safety in an important aspect. The first two conditions of Nozick's knowledge definition jointly state that S truly believes that *p*. Thus, in the context of Nozick's knowledge definition, the sensitivity condition is a counterfactual with a false antecedent. The adherence condition and the safety condition, in contrast, are so-called true-true subjunctives, since theirs antecedents and consequents are both true. DeRose (2004) argues that the truth-conditions for sensitivity

following Stalnaker (1968) and Lewis (1973), has it that we evaluate their truth by looking at possible worlds. Counterfactuals of the form 'If p were the case, then q would be the case' are true according to orthodoxy iff the nearest possible worlds where p is true are such that q is true. Accordingly, we can formulate sensitivity, adherence, and safety in possible worlds terminology as follows:

Sensitivity In the nearest possible worlds where p is false and where S uses M to arrive at a belief whether (or not) p, S does not believe, via M, that p.
Adherence In the nearest possible worlds where p is true and where S uses M to arrive at a belief whether (or not) p, S believes, via M, that p.
Safety In the nearest possible worlds, where S believes that p via M, p is true.[4]

In Sect. 2, I present the problem of knowing necessities for sensitivity and safety accounts of knowledge in more detail. In Sect. 3, I discuss and criticize extant orthodox solutions to this problem as proposed by Nozick (1981) and Pritchard (2009). Section 4 contains a presentation of unorthodox accounts for counterpossibles, involving impossible worlds, as proposed by Nolan (1997) and others. In Sects. 5 and 6, I apply these unorthodox accounts to sensitivity and safety.[5]

2 The necessity problem for sensitivity and safety

Modal knowledge accounts face notorious and well-known problems when it comes to knowledge of necessities.[6] This problem stems from more general problems for counterfactuals involving necessities and impossibilities. Let me briefly sketch how this problem arises. Counterpossibles are counterfactuals with impossible antecedents. Here are two examples:

(CP1) If eight were larger than nine, then I would be three meters tall.
(CP2) If water were H3O, then all textbooks about chemistry would be incorrect.

Footnote 3 continued
are more plausible than those for safety since the meaning and truth conditions of true-true subjunctives are less clear than those of counterfactuals with false antecedents. Moreover, true-true subjunctive conditional face the additional problem that they are trivially true according to the standard counterfactual semantics of Lewis (1973) and Stalnaker (1968). For discussions of this problem and for potential solutions, see McGlynn (2012), Cogburn and Roland (2013), and Walters (2016). In this paper, I will not address these issues concerning true-true subjunctives.

[4] Nozick (1981) and Sosa (1999) prefer subjunctive conditionals, whereas Pritchard (2005 and 2007), a defender of a safety account of knowledge (or at least of a safety-involving account of knowledge) uses possible worlds terminology.

[5] Sensitivity, adherence, and safety are typically discussed as modal conditions for *knowing*. I provide in Melchior (2019) a modal theory of checking arguing that sensitivity is necessary for checking, leaving open whether it is also necessary for knowing. In this paper, I will focus on the necessity problem for modal knowledge accounts.

[6] For a discussion of this problem, see Blome-Tillmann (2017).

(CP1) involves a logical impossibility, (CP2) a metaphysical impossibility.[7] Orthodox semantics has it that a counterfactual is true iff in the nearest possible worlds where the antecedent is true, the consequent is true. Since the antecedents of counterpossibles are impossible, there are no possible worlds where they are true. Hence, according to orthodox semantics, all counterpossibles are trivially or, as Lewis calls it, *vacuously* true.

A similar but more neglected phenomenon also concerns subjunctive conditionals with necessarily true consequents. Here are two examples.

(NC1) If Paris were the capital of France, then 8 would be smaller than 9.
(NC2) If chemistry were fundamentally mistaken, then water would be H_2O.

If the consequent of a counterfactual is true in all possible worlds, then in all possible worlds where the antecedent is true, the consequent is true. Hence, counterfactuals with necessary consequents are also trivially true.[8] Notably, the fact that all counterfactuals with necessary consequents are trivially true is regarded as less worrisome (or it is at least more neglected) than the fact that all counterpossibles are trivially true.[9] However, they are relevant for the purposes of this paper, since the safety condition for beliefs in necessities is a counterfactual of this type.

The fact that counterpossibles and counterfactuals with necessary consequents are trivially true affects modal knowledge conditions. Take sensitivity first. If p is a necessity, then the sensitivity condition 'If p were false and S were to use M to arrive at a belief whether (or not) p, then S wouldn't believe (via M) that p' is a counterpossible.

Hence, every belief in a necessity is trivially sensitive. The safety condition is analogously affected. If p is a necessity then the counterfactual 'If S were to believe that p via M, then p would be true' has a necessary consequent, which is true in all possible worlds, and, therefore, also in all possible worlds where S believes that p via M. Thus, every belief in a necessity is also trivially safe. Notably, there is no impact on the adherence condition. 'If p were true and S were to use M to arrive at a belief whether (or not) p, then S would believe (via M) that p' is non-trivially true or non-trivially false, even if p is a necessity.

These peculiarities have implausible consequences for modal knowledge accounts. Suppose a theory states that S knows that p iff S's belief that p is sensitive and true. In this case, S knows any necessary truth if she believes it. This is counterintuitive since S might come to believe this proposition via an unreliable source, for example via testimony from an unreliable person, or via mere guessing. The same counterintuitive consequences arise for a safety theory of knowledge that states that S knows that p iff S truly and safely believes that p. I will call the

[7] If one rejects the idea that there are metaphysical necessities as defended by Kripke (1980), then only logical impossibilities are relevant.

[8] Here, the notion of vacuousness does not seem to be an adequate metaphor for describing this triviality. Safety, in contrast to sensitivity, is not fulfilled because there is no possible world where the target proposition is false but because it is true in all possible worlds. In order to acquire a unified terminology, I will also say that counterpossibles are trivially true.

[9] This is an interesting fact, given that (NC2) is intuitively false or at least very disturbing.

problem that necessities are trivially known because beliefs trivially fulfil modal conditions the *necessity problem*. Notably, modal knowledge theories do not automatically imply that S knows every necessity believed. The necessity problem arises only if the modal theory contains a claim that sensitivity and/or safety are *sufficient* conditions for converting a true belief into knowledge, being a necessary condition does not suffice to create the problem.[10]

3 Orthodox solutions and their shortcomings

In this section, I will discuss orthodox solutions to the necessity problem and stress their shortcomings in order to motivate an unorthodox solution that also considers impossible worlds. Orthodox solutions try to solve the necessity problem within the framework of orthodox semantics for counterfactuals, i.e. by accepting that counterpossibles and counterfactuals with necessary consequents are trivially true. Let me reflect on two orthodox solutions to the necessity problem and their flaws, the solution proposed by Nozick (1981) and the solutions proposed by Pritchard (2009) and Blome-Tillmann (2017). Nozick (1981, p. 186f) already recognized the necessity problem for his knowledge account. He admits, thereby accepting orthodoxy, that beliefs in necessities automatically fulfill the sensitivity condition. However, he correctly points out that the adherence condition is not automatically fulfilled.

A belief in a necessity violates Nozick's adherence condition if there are many nearby possible worlds where p is true and where S uses M to arrive at a belief whether (or not) p and S does not believe (via M) that p. Suppose that p is a necessity and that S forms the belief that p via mere guessing. There are many nearby possible worlds where S does not believe that p via guessing although p is true. Thus, adherence is not fulfilled. Example: Suppose that S believes via mere guessing truly that 369 + 963 = 1332. It could easily be the case that S did not make this particular guess or made a different guess instead. Hence, there are many nearby possible worlds where 369 + 963 = 1332, where S uses mere guessing, and where S does not believe via mere guessing that 369 + 963 = 1332. Therefore, S's belief fails to fulfill the adherence condition, and S does not know according to Nozick's modal knowledge account. Nozick concludes that for knowing necessities only the truth-condition (1), the belief condition (2), and the adherence condition (4) are necessary and jointly sufficient, but not the sensitivity condition (3).

[10] Nozick explicitly rejects such a problematic view when he assumes that not only sensitivity but also adherence is necessary for knowing. Moreover, Sosa (1999) is also careful in that he only claims that safety is necessary for knowledge, leaving open whether it is also sufficient. Hence, the necessity problem can be avoid if knowledge requires fulfilment of a further condition, one that true beliefs in necessities do not automatically fulfil.

Nozick's account works for the case of mere guessing, but it fails for other cases and, therefore, does not provide a general solution to the necessity problem.[11] S lacks knowledge via M of a necessity *p* according to Nozick's account if there are many nearby possible worlds where S uses M for determining whether *p* is true and where S does not believe that *p* via M. This is the case for mere guessing, since a person might easily believe a different proposition via guessing instead. However, it is contingent on the subject's psychological constitution and on features of the method used whether there are many such nearby possible worlds. Take the following case:

DAMIEN, THE SATANIST
Damien is member of a satanic cult and a poor mathematician. The cult crucially centers on the number 666. A central doctrine of the cult has it that the sum of any two three-digit numbers is 666. Damien has been born into the satanic cult and dogmatically believes its doctrines. Based on this doctrine and due to his mathematical incompetence, he correctly believes that 352+314=666. Moreover, in the nearest possible worlds where 352+314=666 and where Damien consults the doctrines of the cult for determining the sum of 352+314, Damien believes that 352+314=666. Nozick's adherence condition is fulfilled and consequently Damien knows that 352+314=666 according to Nozick.[12]

I think it is a counterintuitive result that Damien knows in this case. Therefore, Nozick's own solution to the necessity problem is not convincing.

The necessity problem not only affects sensitivity, but also safety. Let us have a brief look at an orthodox account that aims to save safety from the necessity problem. Pritchard (2005, 2007) defends an anti-luck epistemology where safety constitutes the required anti-luck condition. Pritchard (2005) originally restricted his safety-based anti-luck epistemology to fully contingent propositions in order to avoid the necessity problem. However, in later writings, Pritchard (2009) extends it to necessities. Originally, for determining whether S's belief that *p* is safe, we look at possible worlds where S believes that *p*. Pritchard (2009, 34) later suggests that for determining safety we look at the whole belief forming process instead of only looking at a particular belief formed via this process. For example, if S believes that $5 + 7 = 12$ via tossing a coin, then there are many nearby possible worlds where this process leads to false beliefs although there are no possible worlds where the particular belief that $5 + 7 = 12$ is false. Hence, S's belief that $5 + 7 = 12$ formed

[11] For an argumentation against Nozick's solution to the necessity problem in the context of checking, see Melchior (2019).

[12] I assume here that the cult crucially relies on the number 666 such that there are no nearby possible worlds where the cult does not teach that the sum of any two three-digit numbers is 666.

Nozick (198, 186f) presents the case of a person S who dogmatically believes a necessity *p* because her parents told her as an example where the adherence condition is violated. However, whether this result can be achieved depends on how we fill in the details. If dogmatically believing that *p* implies that there are no nearby possible worlds where S does not believe that *p*, then adherence *is* fulfilled. In this case, Nozick's example is similar to DAMIEN. Thus, cases of mere guessing provide better examples for adherence violation.

via tossing a coin is safe according to Pritchard's original definition of safety but unsafe according to his revised formulation. Pritchard claims this to be a natural extension of his original safety-based anti-luck epistemology that can perfectly explain why a subject fails to know necessities in such cases.[13]

Blome-Tillmann (2017) discusses the necessity problem for sensitivity and safety. He proposes a similar solution for safety as Pritchard when he suggests replacing safety by the following principle safe': S's belief that *p* (via method M) is safe = df [S couldn't easily have formed a false belief (via M)]. He argues in line with Pritchard that adopting this modified condition can solve the necessity problem for safety. Furthermore, he maintains that there is no analogous solution available for sensitivity. Blome-Tillmann explicitly excludes impossible-worlds accounts for counterfactuals on which I will focus in this paper. He concludes that when it comes to knowing necessities safety is better off than sensitivity.

Prichard's (and Blome-Tillmann's) orthodox solution to the necessity problem for safety faces similar problems as Nozick's solution. Take the following case:

> **RENÉ, THE FERMATIST**
> Suppose that René lives in 1950 and is member of a cult called the Fermatists whose members believe all mathematical theorems that Pierre de Fermat ever proved plus his last theorem. They believe them based on a historical document that just lists these theorems but does not contain any proofs. René believes Fermat's last theorem based on the document, a theorem that has not been proven by 1950. Moreover, all the other propositions that René believes via the document are also necessities. Thus, there is no nearby possible world where René uses the same belief forming method as in the actual world of consulting the document and where the resulting belief (including beliefs of other propositions) is false. Thus, René knows Fermat's last theorem according to Pritchard's revised account.[14]

The outcome that René knows via consulting the document is counterintuitive.[15] Nozick's and Pritchard's orthodox solutions to the necessity problem suffer from similar flaws. Nozick assumes that S lacks knowledge of necessities if there are nearby possible worlds where *p* is true but where S does not belief that *p*. Pritchard argues that S lacks knowledge of necessities if there are many nearby possible worlds where beliefs in other propositions formed via the same method are false. However, the truth of these assumptions are contingent on the modal conditions of the specific cases. Given a specific modal environment, the required modal variation is remote, and S fulfills the required modal condition of adherence or safety although S intuitively does not know.

[13] For an alternative proposal focusing on a priori knowledge, see Miščević (2007).

[14] See also Melchior (2017).

[15] Notably, René's belief of Fermat's last theorem differs from typical cases of testimonial knowledge. For example, one can acquire mathematical knowledge via testimony by reading a textbook that only contains the theorems but not the proofs, but in this case, someone, e.g. the book author, has proven the theorems. However, nobody in the causal chain of René's belief has proven Fermat's last theorem.

Hence, extant solutions to the necessity problems, either proposed by supporters of sensitivity (and adherence) such as Nozick or by supporters of safety such as Pritchard and Blome-Tillmann, are flawed. Both solutions are formulated within an orthodox framework that only considers possible worlds for evaluating counterfactuals. Blome-Tillmann is mistaken when claiming that safety is better off than sensitivity concerning the necessity problem if we restrict ourselves to orthodox solutions. I assume that there is no satisfying solution to the necessity problem for any modal knowledge account within this orthodox framework.

4 Counterfactuals and impossible worlds

Orthodox views about counterfactuals only consider possible worlds for evaluating their truths. They imply that any counterpossible and any counterfactual with a necessary consequent is true and deliver intuitively implausible results. Unorthodox views aim at solving this problem. They allow for consideration of impossible worlds for evaluating counterfactuals.[16] According to these impossible worlds accounts, counterpossibles can turn out to have the truth values that we intuitively attribute to them, i.e. some counterpossibles are true and some are false. In this section, I will present and discuss impossible worlds accounts for counterfactuals, focusing on Nolan's (1997) 'modest approach'.[17] Typically, unorthodox accounts only aim at solving problems for counterpossibles. For the purposes of this paper, I will also reflect on counterfactuals with necessary consequents. In the following sections, I will apply these impossible worlds accounts to sensitivity and safety. Consider the following three counterpossibles, discussed by Nolan (1997):

(CP1) If Hobbes had squared the circle, sick children in the mountains of South America at the time would *not* have cared.
(CP2) If Hobbes had squared the circle, sick children in the mountains of South America at the time would have cared.
(CP3) If Hobbes had squared the circle, then everything would have been the case.

(CP1)–(CP3) are counterpossibles, since squaring the circle is logically impossible. Intuitively, (CP1) is true, but (CP2) and (CP3) are false. However, orthodoxy about counterfactuals delivers the result that (CP1)–CP3) are all true. Orthodoxy does not enable us to treat counterpossibles according to our intuitions.

[16] Orthodox and unorthodox theories about counterfactuals are also labelled vacuism and nonvacuism since orthodoxy has it that all counterpossibles are vacuously true whereas unorthodoxy denies that. See Berto et al. (2018). In this paper, I stick to the terminology of orthodoxy and unorthodoxy since it is intuitively inadequate to call counterfactuals with necessary consequents vacuously true.

[17] There are currently various unorthodox accounts for counterpossibles on the market. Different impossible world accounts will deliver different results for safety and might also deliver different results for sensitivity. I will discuss some results in the following sections. For unorthodox analyses of counterpossibles, see Vander Laan (2004), Brogaard and Salerno (2013), Bjerring (2014), and Berto et al. (2018). For a further impossible world account, see Jago (2013). Tan (2019) defends unorthodoxy for counterpossibles by considering scientific practice and Jenny (2018) by looking at relative computability.

Nolan (1997) begins by suggesting that when evaluating counterfactuals we should also take impossible worlds into account, not only possible worlds, as orthodoxy has it. Accordingly, a counterfactual 'if p were the case, then q would be the case' is true iff in the nearest worlds (possible or impossible) where p is the case, q is also the case. Nolan then argues that impossible worlds are modally and similarity-wise in the same way related to the actual world as possible worlds are. Importantly, different impossible worlds can be differently remote from the actual world. According to Nolan's account, the most remote impossible worlds are the ones where everything is the case, as in (CP3). Nolan calls these worlds *exploding worlds*.

Nolan's impossible-worlds framework enables us to evaluate (CP1)-(CP3) differently. The actual world is such that Hobbes did not square the circle and there are no relevant connections between him and sick children in the mountains of South America at the time. Consequently, impossible worlds where Hobbes squared the circle and where sick children in the mountains of South America at the time did not care about his success are closer to the actual world than impossible worlds where he squared the circle and the children cared and closer than impossible worlds where everything is the case. Counterpossibles are true if in the nearest impossible worlds where p is the case, q is also the case. According to this interpretation, (CP1) is true, but (CP2) and (CP3) are false. This outcome is in line with our pre-theoretical intuitions about these counterpossibles.

Nolan suggests a further restriction to his account about the closeness-relation of possible and impossible worlds that is captured by the following condition:

> **Strangeness of impossibility condition (SIC)**
> Any possible world is more similar (nearer) to the actual world than any impossible world.
> (Nolan 1997, 550)

Since any possible world is closer to the actual world than any impossible world, we consider impossible worlds only when evaluating counterpossibles but not when evaluating counterfactuals with contingent antecedents. Hence, orthodoxy and Nolan's impossible worlds account deliver the same results for counterfactuals with contingent antecedents. In this respect, Nolan regards his impossible worlds account as a conservative, modest, extension of orthodoxy.[18]

Orthodoxy treats counterpossibles and counterfactuals with necessary consequents equally in that they are all trivially true. Nolan's accounts allows for a differentiated evaluation of counterpossibles, but due to SIC, still all counterfactuals with contingent antecedents and necessary consequences are true. Suppose that p is contingent and that q is a necessity. For evaluating the counterfactual 'If p were the case, then q would be the case,' we consider those p-worlds (possible or impossible) that are closest to the actual world. Since p is contingent, there are possible worlds where p is true. Since these possible worlds are closer than any impossible world

[18] Bjerring and Schwarz (2017) object that many impossible worlds accounts cannot be moderate extensions of traditional possible worlds accounts.

according to SIC, we only consider *possible* worlds for evaluating counterfactuals with contingent antecedents. However, given that q is a necessity, q is true in every possible world. Thus, any counterfactual with a contingent antecedent and a necessary consequent is always true on Nolan's account.[19]

This outcome is ensured by SIC. Let us see how rejecting SIC can lead to false counterfactuals with contingent antecedents and necessary consequences. Suppose that p is contingent and that q is a necessity. Suppose further that SIC is false and there are some impossible worlds that are closer to the actual world than some possible worlds. In this case we might also have to consider impossible p-worlds when evaluating 'If p were the case, then q would be the case.' These impossible p-worlds might be such that q is false. In this case, counterfactuals with a contingent antecedent and a necessary consequent can turn out false.

I think that an impossible worlds account for counterpossibles matches our general practice of evaluating counterfactuals by imagining worlds or scenarios that are as similar as possible to the actual world. We stick to this general practice when evaluating counterpossibles, i.e. we consider scenarios involving logical impossibilities keeping as much fixed as possible. For example, when evaluating (CP1)–(CP3), we imagine worlds where squaring the circle is possible, keeping fixed that there is no connection between Hobbes and sick children in the mountains of South America.[20]

Notably, adherents *and* opponents agree that our intuitions about counterpossibles are that some counterpossibles like (CP1) are true, whereas other counterpossibles such as (CP2) and (CP3) are false. Also strict defenders of orthodoxy such as Williamson (2017) acknowledge that we have these intuitions when he aims at explaining them away. Thus, all else being equal, an unorthodox take on counterpossibles that is in line with our intuitions is preferable to an orthodox one that delivers counterintuitive results. I think that unorthodox accounts for counterfactuals face serious challenges but I am optimistic that they can be met.[21]

[19] Nolan (1997) offers SIC as a conjecture about how to think about similarity, but he does not endorse it. In fact, he discusses potential counterexamples against SIC. These are counterfactuals with contingent antecedents and impossible consequences, for example the following assertion of a person who is in awe of Gödel's mathematical ability: If Gödel had believed Fermat's Last Theorem to be false, it would have been. Such counterfactuals might be intuitively true in certain contexts according to Nolan, although they are false, if SIC is true. Notably, statements about safe beliefs in necessities are counterfactuals with contingent antecedents and *necessary* consequents.

[20] For an impossible worlds account of imagination, see Berto (2017).

[21] It has been argued that unorthodoxy about counterfactuals faces serious problems that orthodoxy avoids and that unorthodoxy should be rejected on these grounds. In particular, Williamson (forthcoming) offers a battery of objections against unorthodox interpretations of counterpossibles. His strategy is twofold. First, he points towards problems for unorthodox accounts. For example, he argues that unorthodoxy about counterpossibles implies that the legitimate strategy in mathematics of formulating reductio arguments is mistaken, a result he regards as unsatisfactory. Second, he provides as explanation of our (allegedly) false intuitions that not all counterpossibles are true. Williamson claims that our false intuitions that some counterpossibles are true but that some others are false, rely on bad heuristics. He argues that we take the pair of counterfactuals 'If p were the case, then q would be the case' and 'If p were the case, then *not-q* would be the case' to be contraries such that they cannot both be true. We continue to have this intuition when it comes to counterpossibles and, therefore judge that if one of the 'contrary' counterpossibles is true, then the other must be false. I think that the problems Williamson stresses are

There are currently various impossible worlds accounts for counterfactuals on the market and no canonical version has been established yet. Different accounts, in particular concerning SIC, will imply different takes on the necessity problem for safety and perhaps also for sensitivity. In this paper, I will not develop or adopt a particular account. Rather I will sketch how impossible worlds accounts can be used for solving the necessity problem for sensitivity and safety and address potential challenges for these solutions. In this paper, I will focused on Nolan's (1997) account. Future work on impossible-worlds semantics will presumably deliver a clearer picture of how to precisely solve the necessity problem. I think that there is no convincing solution to the necessity problem available within an orthodox framework. Moreover, I think that impossible-worlds accounts for counterpossibles (and for counterfactuals with necessary consequents) are basically correct. Hence, I think that unorthodox solutions to the necessity problem are on the right track.

5 Impossible worlds for sensitivity

In the next two sections, I will sketch how an impossible worlds account can be used for solving the necessity problem posed for modal theories of knowledge.[22] In this section, I will focus on sensitivity, saving safety for Sect. 6. Recall the sensitivity condition. S's belief that p formed via method M is sensitive iff: If p were false and S were to use M to arrive at a belief whether (or not) p, then S wouldn't believe, via M, that p. If p is necessarily true, then the sensitivity condition for p is a counterpossible. According to Nolan's account, this counterpossible is true, iff in the nearest impossible worlds, where p is false and where S uses M to arrive at a belief whether (or not) p, S does not believe, via M, that p.

By considering examples, we can see that Nolan's account delivers the intuitively correct results for sensitivity, while orthodox accounts deliver intuitively false results. Suppose S uses a perfectly reliable pocket calculator PC_1 for determining the product of 13*14 and the pocket calculary correctly indicates 182. Intuitively, we can come to know mathematical truths by using reliable pocket calculators. S's belief that 13*14 = 182 via using PC_1 is sensitive given an impossible worlds account, because in the nearest impossible worlds where $13*14 \neq 182$ and where PC_1 is used, it does not indicate that 13*14 = 182, since impossible worlds where of $13*14 \neq 182$ and where PC_1 is reliable and consequently not indicating that 13*14 = 182 are closer to the actual world than impossible worlds where $13*14 \neq 182$ and where PC_1 is defective and falsely indicates that

Footnote 21 continued
convincingly rejected by Berto et al. (forthcoming). For example, they show that Williamson's heuristic explanation of our false intuitions about counterpossibles does not generalize and is therefore ad hoc.

[22] For a sketch of an impossible worlds account for a sensitivity-based theory of checking, see Melchior (2019).

13*14 = 182.[23] Thus, S knows that 13*14 = 182 via using PC_1 according to a sensitivity account of knowledge that includes impossible worlds.

Orthodox semantics for counterfactuals also predicts that S knows in this case, since any counterpossible is true according to orthodoxy. The necessity problem for orthodox sensitivity accounts, which do not consider impossible worlds, arises in cases where a subject intuitively does *not* know but where her belief nevertheless turns out to be sensitive. Let us consider such cases. Take the following two examples:

(2)S uses a pocket calculator PC_2 for determining the product of 13*14 that makes random indications.
(3)S uses pocket calculator PC_3 for determining the product of 13*14 that always indicates 182 regardless of what S enters.

Intuitively, neither using PC_2 nor using PC_3 is an appropriate method for determining the product of 13*14. Accordingly, S intuitively does not know in either of these two cases. Take PC_2 first. Suppose PC_2 is completely malfunctioning. Even if PC_2 luckily happens to make an accurate indication, believing based on using PC_2 is not better than luckily making an accurate guess. Thus, a belief formed via PC_2, even if true, does not constitute knowledge. Now take PC_3. PC_3 might not even be a real pocket calculator but a dummy or testing device for eyesight. Likewise, using PC_3 is a flawed method for determining the product of any two numbers.

However, beliefs formed via PC_2 and PC_3 can constitute knowledge according to *orthodox* sensitivity accounts. Suppose that S uses PC_2 for determining the product of 13*14 and PC_2 luckily indicates 182. The corresponding sensitivity condition for S's belief is fulfilled according to orthodox accounts, since it is a counterpossible, which is trivially true. The same holds for a true belief formed via PC_3. In both cases, S knows according to an orthodox sensitivity account that claims that truly and sensitively believing is sufficient for knowing.[24]

Importantly, sensitivity is not fulfilled in these cases according to an impossible worlds account. Among the nearest impossible worlds where $13*14 \neq 182$, there are worlds where PC_2 indicates that 182, since it makes random indications. Hence, S's belief that 13*14 = 182 formed via PC_2 is insensitive, and therefore does not constitute knowledge. Thus, an unorthodox sensitivity account delivers the intuitively correct result in case (2) whereas orthodoxy does not. Let's have a look at case (3). The nearest impossible worlds where $13*14 \neq 182$ are such that PC_3 indicates that 182. This is so because impossible worlds where some arithmetical laws are different but where PC_3 is constructed as in the actual world

[23] In this case, we hold fixed that PC_1 is perfectly reliable and not that PC_1 indicates 182 as the product of 13*14. Take an analogous case for contingent propositions. Suppose that SE is a perfectly reliable search engine for phone numbers. In the nearest possible worlds where S has a different phone number than in the actual world, SE indicates this different phone number for S.

[24] At that point one might suggest adding a further (modal) condition for knowledge that S's belief does not trivially fulfill, but we have already seen that the adherence condition proposed by Nozick is not a proper candidate.

are closer than worlds where these arithmetical laws are different *and* where PC_3 is constructed differently. Thus, S's belief formed via PC_3 is sensitive according to orthodoxy but insensitive according to an impossible worlds account.

An impossible worlds account delivers the desired results for sensitivity accounts of knowledge. If S uses a perfectly reliable pocket calculator, then S knows, since S's belief is sensitive. With pocket calculators that make random indications or always deliver the same indication regardless of what one enters, the resulting belief is insensitive and the subject does not know. These results are in line with the way we approach and evaluate counterfactuals and counterpossibles in generals. We judge whether a counterfactual is true by imagining a world or scenario which is as similar as possible to the actual world except for the fact that 13*14 is not 182 (plus perhaps some arithmetical laws) and imagine what the pocket calculator would indicate in that world. We do not imagine a world where the pocket calculator is constructed differently or worlds that are different from the actual one in every respect like an exploding world.

6 Impossible worlds for safety

The necessity problem not only arises for sensitivity but also for safety. Recall the safety condition: A belief formed via M is safe iff: If S were to believe that *p* via M, then *p* would be true. Impossible worlds accounts provide a differentiated picture for sensitivity in that some counterpossibles are true whereas some others are false. However, as we have already seen, they do not deliver such a differentiated picture for counterfactuals with necessary consequents if the strangeness of impossibility condition, SIC, is accepted. Suppose S believes via M a necessity *p*. Hence, there are possible worlds where S believes via M that *p*. SIC implies that the nearest impossible worlds are more remote than any possible world. Therefore, only *possible* worlds are among the nearest worlds where S believes via M that *p*. Since *p* is true in all possible worlds, S's belief that *p* is safe. Thus, if SIC is true, then any belief in a necessity is safe.

Take the cases of the three different pocket calculators. If S forms a true belief that 13*14 = 182 by using the perfectly reliable pocket calculator PC_1, then her belief is safe because in the nearest worlds (which are only possible worlds) where S believes that 13*14 = 182 via PC_1, the believed proposition is true. Hence, S knows that 13*14 = 182 via PC_1 according to a safety account of knowledge. This is intuitively correct. However, for the same reasons, S's belief that 13*14 = 182 by using PC_2 or by using PC_3 is also safe. Hence, S also knows by using such flawed pocket calculators according to safety theories of knowledge. However, this outcome is intuitively not correct. Thus, an impossible worlds account that accepts SIC has the same counterintuitive consequences for safety as orthodox accounts.

Given SIC, we acquire an unorthodox solution to the necessity problem for sensitivity but still not one for safety. Let's see how an unorthodox safety theory that rejects SIC could handle the necessity problem.[25] Suppose that SIC is false and S believes via M a necessity *p*. S's belief that *p* is not safe iff there are among the nearest worlds where S believes that *p* via M impossible worlds where *p* is false. If one rejects SIC, then one must settle the question of how close impossible worlds can be to the actual world. Settling this issue is a tricky task. However, safety theories of knowledge provide at least information about how close impossible worlds must be such that a belief in a necessity can fail to be safe. A crucial motivation for safety theories is to provide a solution to the skeptical problem. Safety theorists such as Sosa (1999) and Pritchard (2005, 2007) prefer a Moorean solution to the skeptical problem according to which we know that the skeptical hypothesis is false. They argue that our beliefs that we are not brains in vat are trivially safe since any world where we are brains in vat is very remote.[26] Hence, for determining whether a belief is safe, we only consider possible worlds that are closer to the actual world than worlds where we are brains in vats.[27] Accordingly, any belief in a necessity is trivially safe if any impossible world is at least as remote as possible worlds where we are brains in vats. In this case, rejecting SIC does not solve the necessity problem for safety.

Suppose for the sake of argument that there are impossible worlds that are sufficiently close to the actual world. In particular, suppose that there are many nearby impossible worlds where 13*14 = 182 is false. S's belief that 13*14 = 182 is safe iff in the nearest worlds where S believes that 13*14 = 182 via method M it is true that 13*14 = 182. Suppose S uses PC_1, a perfectly reliable pocket calculator that correctly indicates that the product of 13*14 is 182. In the *possible* worlds where PC_1 is used, it correctly indicates that the product of 13*14 is 182. In the *impossible* worlds where PC_1 is used and where the product of 13*14 is not 182 it correctly indicates something else. Hence, there are no nearby worlds where PC_1 falsely indicates that 13*14 = 182. Consequently, S safely believes that 13*14 = 182 via PC_1. Suppose that S uses PC_2 that makes random indications. In this case, there are among those worlds where PC_2 indicates that 13*14 = 182 (sufficiently many) impossible worlds where $13*14 \neq 182$ and S's belief is unsafe.

[25] Some defenders of impossible worlds accounts defend SIC at least on theoretical grounds, but the overall verdict is not clear. Mares (1997) claims that the view that all possible worlds are closer than any impossible world seems reasonable. Also Bjerring (2014) accepts a version of SIC. Nolan (1997, 550) suggests that SIC has a 'fair bit of intuitive support'. Nevertheless, he offers SIC only as a conjecture about how we treat relative similarity and admits that there might be some exceptions. However, the exceptions that he discusses are counterfactuals with contingent antecedents and *impossible* consequents. For a defense of SIC against these counterexamples, see Berto and Jago (2019). Vander Laan (2004) argues that conversational considerations suggest that impossible worlds sometimes are, in relevant respect, closer to the actual world than some possible worlds. Berto (2013) expresses the intuition, that some impossible worlds can be closer than some possible worlds, but without explicitly arguing for it.

[26] Moreover, they are committed to assuming that other skeptical scenarios such as being deceived by an evil demon are at least equally remote as being a brain in a vat.

[27] While for determining sensitivity the neighborhood of possible worlds varies with the proposition believed, for determining safety it remains the same modal neighborhood for every proposition. See Zalabardo (2017).

Suppose now that S uses PC_3 that always indicates 182 regardless of what one enters because PC_3 is constructed in way such that it does not easily make an indication other than 182. Again in this case, there are among the worlds where S believes that 13*14 = 182 via PC_3 impossible worlds where this equation is false and S's belief is unsafe.

If we reject SIC and assume that there are sufficiently many impossible worlds where 13*14 = 182 is false among the nearby worlds, i.e. sufficiently many impossible worlds are sufficiently close, unorthodox safety accounts deliver the desired result. S can know via a perfectly reliable pocket calculator PC_1 but knows neither via PC_2 nor via PC_3. Without making these two additional assumptions, such a solution is not available.

We have already noted, that defenders of impossible world accounts do not agree about whether SIC is true. On the one hand, rejecting SIC has a certain intuitive appeal. For example, it seems plausible to accept that, all else being equal, impossible worlds where some technical logical details are different are closer than worlds where I am a brain in vat or the only existing human being. Moreover, considering impossible worlds and possible worlds is presumably in line with our practice of imagination, e.g. when we consider impossible worlds for evaluating whether S's belief formed via PC_2 or PC_3 is safe.

On the other hand, impossible worlds accounts that reject SIC face serious challenges. First, on these accounts, we must also consider impossible worlds for evaluating counterfactuals with *contingent* antecedents. Hence, these accounts will deliver other results than orthodoxy for counterfactuals with contingent antecedents and are in this respect non-conservative extensions of the orthodox semantics for counterfactuals. However, how to evaluate counterfactuals with contingent antecedents by considering impossible worlds is an open question.[28] Second, if impossible worlds can be closer to the actual world than possible worlds, then the question comes up which impossible worlds can be. Can only *metaphysically* impossible worlds be closer, or also *logically* impossible worlds? Can only logically impossible worlds where some logical details are different be closer, or also impossible worlds where the most fundamental logical laws do not hold? These are serious questions that have to be settled if one opts for an impossible worlds account that declines SIC. Thus, many might reject such a theory on theoretical grounds. In this case, impossible worlds can offer a solution to the necessity problem for sensitivity but not for safety. Otherwise, a solution for safety is also available. As it stands sensitivity theories are better off than safety theories, against what adherents of safety such as Blome-Tillmann suggest.

[28] Bjerring and Schwarz (2017) argue that many impossible worlds accounts are not conservative extensions of traditional possible worlds accounts. If this is correct then sensitivity might not be automatically better off than safety when it comes to solving the necessity problem.

7 Conclusion

The necessity problem relies on orthodox semantics for counterfactuals according to which every counterpossible and every counterfactual with a necessary consequent is trivially true. This problem arises both for sensitivity and safety accounts of knowledge. Orthodox solutions to the necessity problem as proposed by Nozick (1981) and Pritchard (2009) are unsatisfactory. A moderate impossible worlds account that accepts SIC, as defended by Nolan (1997), delivers the intuitively correct result that some beliefs in necessities are sensitive and can, therefore, constitute knowledge whereas others are not. However, SIC prevents us from acquiring an analogous result for safety. S's belief in a necessity *p* can only turn out to be unsafe if SIC is rejected. One might regard the resulting non-conservative impossible worlds account as rather eccentric and coming at too high a cost. As for sensitivity, a conservative unorthodox extension of possible worlds accounts can solve the necessity problem. As for safety, either we have to accept a non-conservative unorthodox extension or the necessity problem remains unsolved.

Acknowledgements Open access funding provided by University of Graz. An earlier version of this paper was presented at the *2018 Metaphysics Workshop* at the IUC Dubrovnik. I am thankful to the audience for the discussion and to Wes Siscoe for comments on this paper. The research was funded by the Austrian Science Fund (FWF): P 28884-G24.

References

Adams, F., & Clarke, M. (2005). Resurrecting the tracking theories. *Australasian Journal of Philosophy, 83*(2), 207–221.

Berto, F. (2013). Impossible worlds. In Edward N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2013 Edition). Retrieved from https://plato.stanford.edu/archives/win2013/entries/impossible-worlds/.

Berto, F. (2017). Impossible worlds and the logic of imagination. *Erkenntnis, 82*(6), 1277–1297.

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

Berto, F., et al. (2018). Williamson on counterpossibles. *Journal of Philosophical Logic, 47*, 693–713.

Bjerring, J. C. (2014). On counterpossibles. *Philosophical Studies, 168*(2), 327–353.

Bjerring, J. C., & Schwarz, W. (2017). Granularity problems. *Philosophical Quarterly, 67*(266), 22–37.

Blome-Tillmann, M. (2017). Sensitivity actually. *Philosophy and Phenomenological Research, 94*(3), 606–625.

Brogaard, B., & Salerno, J. (2013). Remarks on counterpossibles. *Synthese, 190*, 639–660.

Cogburn, J., & Roland, J. W. (2013). Safety and the true–true problem. *Pacific Philosophical Quarterly, 94*(2), 246–267.

DeRose, K. (2004). Sosa, safety, sensitivity, and skeptical hypotheses. In J. Greco (Ed.), *Ernest Sosa and his critics* (pp. 22–41). Malden, MA: Wiley Blackwell.
Jago, M. (2013). Impossible worlds. *Noûs, 47*(3), 713–728.
Jenny, M. (2018). Counterpossibles in science: the case of relative computability. *Noûs, 52*(3), 530–560.
Kripke, S. A. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
Kripke, S. A. (2011). *Nozick on knowledge. Philosophical Troubles. Collected papers* (Vol. 1, pp. 162–224). Oxford: Oxford University Press.
Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
Mares, E. D. (1997). Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic, 38*, 516–526.
McGlynn, A. (2012). The problem of true-true counterfactuals. *Analysis, 72*(2), 276–285.
Melchior, G. (2017). Epistemic luck and logical necessities: armchair luck revisited. In S. Borstner & S. Gartner (Eds.), *Thought experiments between nature and society: A Festschrift for Nenad Miščević* (pp. 137–150). Cambridge: Cambridge Scholars Publishing.
Melchior, G. (2019). *Knowing and checking. An epistemological investigation*. New York: Routledge.
Miščević, N. (2007). Armchair luck: Apriority, intellection and epistemic luck. *Acta Analytica, 22*(1), 48–73.
Nolan, D. (1997). Impossible worlds: a modest approach. *Notre Dame Journal for Formal Logic, 38*, 535–572.
Nozick, R. (1981). *Philosophical explanations*. Cambridge, Mass.: Harvard University Press.
Pritchard, D. (2005). *Epistemic luck*. Oxford: Oxford University Press.
Pritchard, D. (2007). Anti-luck epistemology. *Synthese, 158*, 277–297.
Pritchard, D. (2009). Safety-based epistemology: whither now? *Journal of Philosophical Research, 34*, 33–45.
Sosa, E. (1999). How to defeat opposition to Moore. *Philosophical Perspectives, 13*, 141–153.
Stalnaker, R. (1968). A theory of conditionals. *Studies in logical theory, American Philosophical Quarterly Monograph Series, 2* (pp. 98–112). Oxford: Blackwell.
Tan, P. (2019). Counterpossible non-vacuity in scientific practice. *Journal of Philosophy, 116*(1), 32–60.
Vogel, J. (1987). Tracking, closure and inductive knowledge. The possibility of knowledge. In S. Luper-Foy (Ed.), *Nozick and his critics* (pp. 197–215). Totowa, NJ: Rowman and Littlefield.
Vander Laan, D. (2004). Counterpossibles and similarities. In F. Jackson & G. Priest (Eds.), *Lewisian themes: The philosophy of David K. Lewis* (pp. 258–275). Oxford: Clarendon Press.
Walters, L. (2016). Possible world semantics and true-true counterfactuals. *Pacific Philosophical Quarterly, 97*(3), 322–346.
Williamson, T. (2017). Counterpossibles in metaphysics. In B. Armour-Garb & F. Kroon (Eds.), *Philosophical fictionalism*. Oxford: Oxford University Press.
Zalabardo, J. L. (2017). Safety, sensitivity and differential support. *Synthese*. https://doi.org/10.1007/s11229-017-1645-z.