

## ARTICLE

# Finding Closure for Safety

Moritz Schulz 

Universität Hamburg, Germany  
Email: [moritz.schulz@uni-hamburg.de](mailto:moritz.schulz@uni-hamburg.de)

(Received 30 January 2019; revised 25 November 2019; accepted 11 December 2019;  
first published online 18 February 2020)

## Abstract

There are two plausible constraints on knowledge: (i) knowledge is closed under competent deduction; and (ii) knowledge answers to a safety condition. However, various authors, including Kvanvig (2004), Murphy (2005, 2006) and Alspecter-Kelly (2011), argue that beliefs competently deduced from knowledge can sometimes fail to be safe. This paper responds that one can uphold (i) and (ii) by relativizing safety to methods and argues further that in order to do so, methods should be individuated externally.

**Keywords:** knowledge; safety; closure; methods; externalism

## 1. Safety and closure

Perhaps the biggest advantage a safety condition on knowledge has over its main rival, a sensitivity condition, is that it promises to explain, or at least not to violate, the closure of knowledge under competent deduction.<sup>1</sup> Although some philosophers, most prominently Dretske (1970, 2005) and Nozick (1981: 207f.), think that a failure of closure might help in solving the problem of skepticism, many would not be ready to give up on a closure principle for knowledge.<sup>2</sup> From a closure-friendly perspective, safety should be favored over sensitivity because it does not conflict with closure.

A salient way of understanding closure is the following (cf. Williamson 2000: 117 and Hawthorne 2005: 29):

**Closure.** If a subject (i) knows  $P_1, \dots, P_n$ , (ii) competently deduces  $Q$  from  $P_1, \dots, P_n$  (in favorable circumstances) and (iii) thereby comes to believe  $Q$ , then the subject knows  $Q$ .

This closure principle expresses the plausible thought that knowledge can be extended by deduction. If one knows a number of premises  $P_1, \dots, P_n$  and competently deduces from them a conclusion  $Q$  and thereby comes to believe  $Q$ , then one knows  $Q$ .<sup>3</sup> The

<sup>1</sup>But see Bernecker (2012) and Kvanvig (2004) for the view that sensitivity and safety are actually on a par with respect to closure.

<sup>2</sup>See e.g. Hawthorne (2005), Pritchard (2005: 167f.), Sosa (1999) and Williamson (2000: 117).

<sup>3</sup>The present version of a closure principle for knowledge differs from some variants in the literature which require knowledge of the validity of the inference instead of competent deduction (see Bernecker 2012: Section 5). As far as I can see, this difference is insubstantial for the main arguments in this paper.

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

bracketed qualification ‘in favorable circumstances’ is meant to exclude cases where one loses one’s knowledge of the premises during the reasoning process or is exposed with a defeater for the correctness of one’s reasoning at the exact moment when one reaches the conclusion (see Kvanvig 2006 for discussion). I mention these possibilities only to set them aside in what follows.

With this in mind, let us turn now to safety as introduced by Pritchard (2005), Sosa (1999), Williamson (2000: Ch. 5) and others. A safety condition on knowledge is just this (various ways of explicating safety will be discussed below):

**Safety Condition.** If a subject knows  $P$ , then the subject’s belief in  $P$  is safe.

Here I introduce safety merely as a necessary condition on knowledge. There is work on strengthening the concept of safety, so that it might have a claim to be both necessary and sufficient for knowledge (see Grundmann 2018).

What makes a safety condition seem closure-friendly? A key feature of safety is that it is about what is true at a certain range of close worlds. And if a number of premises are true across a set of worlds, then so too is anything which logically follows from these premises. This shows that a prototypical safety condition, which I’ll label ‘Naïve Safety’, is closed under logical consequence:

**Naïve Safety.** A subject’s belief in  $p$  is safe iff  $p$  is true in all close worlds.

If **Naïve Safety** were a plausible condition on knowledge, it would pose no threat to a closure principle. **Naïve Safety** is closed under logical consequence in the sense that if a number of beliefs are safe, then any belief which can be deduced from these beliefs is safe as well. So, any belief competently deduced from known premises would be safe, for competent deduction arguably requires that the conclusion follows from the premises.

The problem with **Naïve Safety** is that it is not a plausible condition on knowledge. A case which shows this is Nozick’s (1981: 179) grandmother example. As a crucial feature of Nozick’s original case is not necessary to argue against **Naïve Safety**, the following story describes a simplified version which I relabel to mark the difference:

#### *Grandfather*

A grandfather goes to a hospital to visit his granddaughter. She is very sick and could die any minute. But as a matter of fact, when the grandfather enters the room, his granddaughter is still alive. He sees her and forms the belief that his granddaughter is alive.

The grandfather knows in this case that his granddaughter is alive. However, given the circumstances, she could easily have been dead. Consequently, it is not true in all close worlds that she is alive. If so, safety as understood along the lines of **Naïve Safety** is not necessary for knowledge.

For this and various other reasons, current defenders of a safety condition propose more nuanced versions of safety. The issue with more nuanced versions of safety is that the easy argument in favor of safety’s closure under logical consequence no longer holds. More problematically, a number of potential counterexamples have been designed to show that knowledge would not be closed under competent deduction on more refined safety conditions (Kvanvig 2004; Murphy 2005, 2006; Alspector-Kelly 2011). My aim in this paper is to show that this is not correct. There are ways of articulating a safety condition which pose no threat to the closure of knowledge under competent deduction. On the upside, this means that one can coherently

hold that knowledge answers to a safety condition and that knowledge is closed under competent deduction. As it turns out, the downside of this is that deduction faces the same kind of intricate issues of individuation as perception and other belief-forming methods.

The paper proceeds as follows. Section 2 reviews potential counterexamples. Section 3 introduces methods. Section 4 discusses the possibility that inferential methods are partly individuated by the premises to which they are applied. Section 5 offers a way of individuating inferential methods externally and shows that on such a view, safety poses no obstacle to knowledge being closed under competent deduction. Section 6 discusses a residual issue and section 7 concludes.

## 2. Problem cases

In response to problems with **Naïve Safety**, the notion of safety has been refined in various respects. Perhaps the most important change is to focus not merely on whether the proposition is true in close worlds but rather on whether the proposition is true if the subject believes it to be (Sosa 1999; Williamson 2000: Ch. 5; Pritchard 2005):

**Safety (belief version).** A subject's belief in  $P$  is safe iff  $P$  is true in all close worlds in which the subject believes  $P$ .

This gets around the problem of known propositions which could easily have been false. To wit, recall the grandfather case. Although what the grandfather believes, namely that his granddaughter is alive, could easily have been false, his belief is modally robust. He would not have believed that his granddaughter is alive in circumstances in which she had been dead, for he would then not have seen her in her room.

The most prominent counterexamples to safety's closure are directed at the belief version of safety. Consider this case presented by Murphy (2005):<sup>4</sup>

### *Probabilistic barn*

Harriet looks out of her car window and sees a *red barn* in the field. She concludes, by deduction, that there is a barn in the field. As a matter of fact, there is a red barn in the field, but there could have easily been, with 99% probability, a *green barn facade* in the field. (Let's suppose the mayor previously held a lottery on whether to put a red barn or a green barn facade in the field.)

Given that, as a matter of fact, everything is normal – there is a red barn in the field, lighting conditions are good and there are no barn facades nearby – Murphy (2005) is right, I think, in assuming that Harriet knows

- (1) There is a red barn in the field.

If knowledge is closed under competent deduction, it should be possible for Harriet to gain deductive knowledge of (2) on this basis:

- (2) There is a barn in the field.

<sup>4</sup>Obviously, this case is a variation of Goldman's (1976: 773) fake barn scenario. It also seems inspired by Kripke's discussion of this case in his lectures on Nozick in Princeton. Kvanvig (2004: 209) offers a similar counterexample to closure. See Pritchard (2005: 167f.) for a response to Kvanvig. Murphy (2006: 372) presents a further case with a similar underlying structure as the one discussed here.

Note that ascribing deductive knowledge of (2) is very plausible: we would credit Harriet with knowledge that there is a barn in the field if she deduced this from her knowledge that there is a red barn in the field.

But now observe that Harriet's belief in (2) does not appear to be safe. In some close worlds, the mayor would have put up a green barn facade. Harriet would have believed that there is a green barn in front of her from which she would then have inferred (2) by way of the same inference rule as in the actual world. But (2) would have been false, for there would only be a barn facade in the field. For this reason, Murphy (2005) concludes that Harriet's belief in (2) is not safe.

It is not fully clear whether one should follow Murphy in his diagnosis of this case. One way to resist his argument would be to deny that worlds in which barn facades are put up are close. Despite the high antecedent probability of a barn facade having been put up, such a maneuver would not necessarily be ad hoc, for safety is a modal rather than a probabilistic notion (Pritchard 2014).<sup>5</sup>

But even if one finds fault with the details of Murphy's case, the structure behind it reveals a recipe for constructing interesting problem cases. Take any safe belief  $P_1$  from which  $Q$  is competently deduced. Given that logical consequences are often weaker than the premise from which they are inferred, there are plenty of cases where  $Q$  could also have been inferred from a different premise  $P_2$  which is logically independent of  $P_1$ . Now see to it that there are close worlds in which  $P_1$ ,  $P_2$  and  $Q$  are false but the subject believes  $P_2$  and infers  $Q$  from it in the same way as in the actual world. Then  $Q$  is not true in all close worlds in which it is believed and so safety would fail.

Alspector-Kelly (2011) presents a similar case which instantiates this recipe. It goes like this (slightly adapted):

#### *Parking lot*

Luke has parked his very expensive car in parking lot B1. The parking garage comes with an app, *smart lot*, which allows Luke to monitor B1 on his phone. Later that night Luke checks his app (out of curiosity, we may suppose), sees his car in B1 and concludes that it has not been stolen. Unbeknownst to him the attendant has just started to steal a car every night (it is a very big garage). However, had Luke seen on his app that his car was no longer in B1, he would have thought that the attendant had moved it to another lot (which is something the attendant sometimes does). Luke would still have believed that it has not been stolen.

It is evident why this is another problem for closure. Luke's belief that his car is in B1 may very well be knowledge. Seeing his car on a well-functioning app is almost as good as direct perception. Although there are close worlds in which his car is not in B1 (either because it was stolen or because it was moved to another lot), these are worlds in which he does not believe that it is in B1 because he would not see it being in B1 on his app. However, in some of these worlds, Luke still believes his car has not been stolen despite the fact that it has. In such worlds, he concludes that it has not been stolen from it being in another lot which is something he believes in those worlds in which the car is not in B1. Again, despite knowing that his car is in B1, the conclusion that it has not been stolen does not appear to be safe.

I find it hard to object to Alspector-Kelly's example by deeming the problematic worlds not to be close. After all, some cars are moved every night and one car even gets stolen. If Luke's car could not be said to have 'easily been moved' and 'easily been stolen', it would be hard to keep a firm grip on the intended sense of 'close world'.

<sup>5</sup>See also the response by Williamson (2009b) to Hawthorne and Lasonen-Aarnio (2009).

If we grant the counterexamples (or at least the possibility of instantiating the recipe for constructing counterexamples mentioned above), what are our options? The most radical option would be to conclude that knowledge is not closed under competent deduction. But recall that this conclusion is not supported by the problematic cases. To the extent that it is plausible that the relevant premise is known in these cases, it is equally plausible that the conclusion is known. Harriet knows that there is a barn in the field when inferring this from her knowledge of a red barn being in the field. And Luke knows that his car has not been stolen when inferring this from it being in B1. Therefore, the examples are best seen as challenging whether safety is necessary for knowledge. If the conclusion of the inference in the two cases is known but not safe, then safety is not necessary for knowledge.

I think this might be the right conclusion to draw as long as safety is understood along the lines of **Safety (belief version)**. But independently of issues with closure, the debate on safety indicates that **Safety (belief version)** might not be the best way of explicating a plausible safety condition anyway. If this is so, then one should first look at safety conditions independently offered as improvements of **Safety (belief version)** before concluding that safety is at odds with knowledge's closure. This is what I'll try to do next.

### 3. Methods

Nozick's (1981: 179) original grandmother case illustrates why safety conditions need an increased complexity:

#### *Grandmother*

A grandmother goes to a hospital to visit her grandson. Her grandson is very sick and could die any minute. But as a matter of fact, when the grandmother enters the room, her grandson is still alive. She sees him and forms the belief that her grandson is alive. However, had the grandson died, her relatives would have told the grandmother that her grandson is still alive and not let her see him.

It seems intuitively true that the grandmother knows in this case that her grandson is alive. However, given the circumstances, there are close worlds in which the grandmother falsely believes that her grandson is alive because her relatives lie to her about this. **Safety (belief version)** does not seem necessary for knowledge.

Close worlds in which the grandmother falsely believes that her grandson is alive differ in an interesting respect from the actual situation. The grandmother uses a different *belief-forming method*. Her belief is based on testimony and not on perception. In the actual world, she believes what she believes because of the facts represented by her perceptual state (her grandson lying in his bed saying "Hello!"), while in the counterfactual scenario her belief is based on what her relatives say ("Your grandson is alive but can't talk right now – let's go to the cafeteria!").

This observation offers a way of improving a safety condition. Instead of **Safety (belief version)**, one may propose the following account of safety (see e.g. Pritchard 2005: 156):

**Safety (method version)**. A subject's belief in  $p$  is safe iff  $p$  is true in all close worlds in which the subject believes  $p$  by using the same method as in the actual world.<sup>6</sup>

<sup>6</sup>Nozick (1981: 180) discusses the possibility of a belief being formed by more than one method, testimony plus perception, for example. We may either set such cases aside for now or treat cases of multiple methods as cases of *compound methods* having other methods like perception or testimony as parts.

Does this version of safety help with closure? As it turns out, this crucially relies on what counts as *the same method*. In the upcoming two sections, I therefore look at two potential constraints on the individuation of methods which would make the closure of knowledge under competent deduction consistent with a safety condition on knowledge.

#### 4. Methods and premises

One way of understanding an inferential method essentially involves the premises which it takes as input. On this thought, one does not use the same method if the premises of one's reasoning differ. This would be so even if there is no difference in the inference rules applied. Thus, when evaluating whether a given belief is safe, one would look only at close worlds in which this belief is based on the same premises. This gives rise to the following constraint on the individuation of methods:

**Inferential Methods (grounds version).** Two beliefs are formed by the same deductive method only if they are inferred from beliefs with the same propositional content (only if they are inferred from the same premises).

This constraint specifies a necessary condition on the individuation of methods. It allows one to distinguish two deductive methods which start out from different premises. The constraint is not meant to fully individuate deductive methods.

It is easy to see that **Safety (method version)** can circumvent the counterexamples against closure when inferential methods are construed as involving the premises essentially. Start with Murphy's barn case. When Harriet falsely believes in a close world that there is a barn in the field, she infers this from the belief that there is a green barn in the field. In the actual world, however, she has a different ground for her belief. There she believes this because she believes that there is a red barn in the field.

A similar consideration applies to Alspecter-Kelly's parking lot case. In the actual world, the belief that the car is not stolen is inferred from it being in B1, while in the counterfactual scenario, the belief is inferred from a different premise, namely the belief that the attendant has moved it to a different lot.

Alspecter-Kelly (2011: 136f.) objects to a strategy of this kind with a modified counterexample. The idea behind the modified counterexample is simple: just as one may believe the conclusion of an argument for different reasons in close worlds, so one could also believe the premises for different reasons. If this makes the premises false in some close worlds, then the conclusion may be false as well despite being believed for the same reason as in the actual world. Here is the modified counterexample:

##### *Parking Lot (modified)*

Luke has parked his very expensive car in parking lot B1. The parking garage comes with an app, *smart lot*, which allows Luke to monitor B1 on his phone. Later that night Luke checks his app (out of curiosity, we may suppose), sees his car in B1 and concludes that it has not been stolen. Unbeknownst to him the attendant has just started to steal a car every night (it is a very big garage). *Modification:* Luke is disposed to call the attendant if he does not see on his app that the car is in B1. Moreover, the attendant would cut the wires to Luke's app if the attendant decides to steal Luke's car. However, when Luke would call him he would tell Luke that the car is still in B1. Trusting the attendant, Luke would believe that his car has not been stolen based on his belief that it is still in B1.

Take one of the worlds in which Luke's car is stolen. In such a world, Luke believes by testimony that his car is (still) in B1 and infers from the latter that it has not been stolen (which is false). So, there are close worlds in which Luke's inferred belief is false despite being based on the same premise as in the actual world.

Given that in the actual world, Luke sees his car in B1 via a well-functioning app, it is hard to deny that he knows that his car is in B1. Although he would falsely believe that his car is in B1 in a close world, this belief would be based not on what the app shows him but rather on a different method, namely on testimony (what the attendant tells him). This aspect of the case is structurally similar to Nozick's grandmother case. Again, it also seems that the inferred belief – that his car has not been stolen – constitutes knowledge. As far as what the case *prima facie* suggests, we are not looking at a problem for closure but rather at a problem for safety.

From the perspective of a defender of **Safety (method version)** who individuates deductive methods partly by their premises, a fairly natural response would be to draw a distinction between *direct premises* and *indirect premises*. A direct premise is the penultimate step in a chain of reasoning. Indirect premises are all beliefs which occur at an earlier stage of the reasoning as intermediate conclusions or as ultimate, non-inferential grounds the reasoning is based on. A *premise* (or *ground*) of a belief could then be construed either in a narrow sense meaning *direct premise* or in a broader sense meaning *direct or indirect premise*.

In a similar vein, an inferential method can be construed in a *narrow* sense where the methods just cover the process of inferring the conclusion from the direct premises it is based on. Or, alternatively, methods can be construed *broadly* to cover the whole belief-forming process with all its intermediate steps. On this approach, the method a subject uses would even include the way the most fundamental beliefs are formed. For this reason, there are broad methods which are hybrid in the sense of being neither purely deductive nor purely non-deductive. For instance, a base belief may be formed non-deductively, by perception say, and then a further belief is formed by deduction. The broad method for the latter is then a chain consisting of a perceptual and an inferential link.

The distinction offers an alternative understanding of safety:

**Safety (broad method version).** A subject's belief in *p* is safe iff *p* is true in all close worlds in which the subject believes *p* by using the same *broad method* as in the actual world.

One would pair this understanding of safety with a revised version of how to individuate inferential methods:

**Inferential Methods (broad grounds version).** Broad belief-forming methods A and B are the same only if A and B take as input the same direct and indirect premises or grounds.

Alspector-Kelly's adjusted counterexample would work on a narrow sense of 'method', for the direct reason – that the car is in B1 – is kept constant over the relevant class of close worlds. Interestingly, however, the counterexample would not challenge safety on the revised account. In those close worlds in which Luke falsely believes that his car has not been stolen based on his false belief that the car is in B1, Luke believes the latter on different grounds than in the actual world. In the actual world he believes that his car has not been stolen because he sees his car in B1 on his app, while in the relevant counterfactual world he believes that his car is in B1 because

the attendant tells him so. The problematic worlds are therefore worlds in which the broad method differs.

Although this protects the present approach against Alsdpector-Kelly's modified counterexample, is there any guarantee that a different counterexample could not easily be constructed? In other words: Can one devise a positive argument that **Safety (broad method version)** is consistent with closure under competent deduction when methods are partly individuated according to **Inferential Methods (broad grounds version)**?

This question has a positive answer. The principle one can show to follow is this:

**Safe Deduction.** If a subject (i) knows  $P_1, \dots, P_n$  and (ii) competently deduces  $Q$  from  $P_1, \dots, P_n$  and (iii) believes  $Q$  because of (i) and (ii), then the subject's belief in  $Q$  is safe.

The principle states that beliefs competently deduced from known premises are safe.<sup>7</sup> Counterexamples of the kind proposed by Murphy and Alsdpector-Kelly could then not be found. For any time there is a case of an inferential belief based on known premises, **Safe Deduction** would imply that the inferred belief is safe.

With the present constraints on safety in place, **Safe Deduction** can be established in the following way. Take the premises  $P_1, \dots, P_n$  to be known and assume that  $Q$  is competently deduced from them. Is the subject's belief in  $Q$  safe? Let  $W$  be the set of close worlds in which the subject uses the same broad method. By **Inferential Methods (broad grounds version)**, one can conclude that  $W$  includes only worlds in which  $Q$  is inferred from  $P_1, \dots, P_n$ . Moreover, the broad method the subject uses for inferring  $Q$  has as parts the broad methods the subject has used to arrive at  $P_1, \dots, P_n$ . Therefore, the set  $W$  of worlds in which the subject believes  $Q$  by way of the same broad method is a subset of the set  $W'$  of worlds in which  $P_1, \dots, P_n$  are believed by way of the same broad method as in the actual world. Given that  $P_1, \dots, P_n$  are known in the actual world, they are, by assumption, safe (here one makes use of **Safety Condition**). For this reason, they are true in all worlds in  $W'$  and therefore, given  $W \subseteq W'$ , in all worlds in  $W$ . Now add to this that a deduction is only competently made if the conclusion follows from the premises (see the upcoming section for more on this) and one sees that  $Q$  must be true in  $W$ . This shows that  $Q$  is safe in the sense of **Safety (broad method version)** as desired.

These findings comprise the good news. If safety is constrained by the method one uses and if methods are understood broadly to pertain to all steps of the way leading up to the target belief and if they are, moreover, constrained by the content of the premises involved, then safety poses no threat to knowledge's closure.

The envisaged account of methods individuates methods very finely. Inferential methods  $A$  and  $B$  can only be the same if they represent exactly the same "inferential tree": all direct and indirect premises must be the same. This makes it very hard for the same method to be applied by two persons or by the same person more than once.<sup>8</sup> Moreover, this way of individuating methods is at odds with a plausible pre-theoretical constraint on methods. A method should be applicable to a variety of input states.<sup>9</sup> But on the present account, a given method can only be applied to the same premises or grounds. The only variety a method would allow for are different tokens of the same premises. This may not be a knock-down objection, for in a theory of knowledge

<sup>7</sup>To clarify, this principle does not express that safety is closed under competent deduction. This would be a slightly stronger claim. However, it is easily seen that the argument below can be adapted to verify this stronger claim. For discussion, see Murphy (2006).

<sup>8</sup>Many thanks to an anonymous referee for alerting me to this problem.

<sup>9</sup>See Nozick (1981: 233) on the *generality* a method should have.

'method' is partly a technical term coined for representing ways in which a belief is formed.<sup>10</sup> Nevertheless, even if construed in a technical sense, extremely finely individuated methods reduce the logical strength of a safety condition. As the individuation of methods gets finer and finer, there are fewer and fewer (close) worlds in which beliefs are formed by the same method. Hence, there are fewer possibilities for a subject's belief to violate safety. In other words, individuating methods more finely makes it easier for a belief to be safe. If this is pushed to the extreme, safety becomes a fairly weak condition on knowledge in danger of losing much of its explanatory power. For this reason, it is worthwhile to see whether one cannot find a way of defending closure which provides methods with a broader realm of possible application.

### 5. External individuation of methods

A different defense strategy can be used for making safety compatible with knowledge's closure. It exploits the fact that the debate on safety indicates that methods must be individuated externally for reasons independent of issues with closure (Broncano-Berrocal 2014; Grundmann 2018). In light of this, it seems reasonable to see what an external individuation could mean for deduction.

For starters, let us explicitly distinguish *deduction as mere reasoning* from *deduction as a belief-forming mechanism*. Deduction may take as its starting point any kind of assumption. Perhaps an assumption is merely made for the sake of argument. Perhaps an assumption is considered possible and one is interested in what would follow from it should it turn out to be true. Whatever conclusion follows in such cases, the reasoning does not usually produce a belief in the conclusion, for one does not believe the premises one started with. Still, the reasoning employed may look the same as in a case in which one does believe the premises and therefore comes to believe the conclusion. The same inference rules may be applied to the same propositions in the same order. So, there is a sense of deduction where it refers to deductive reasoning which does not by itself constrain which epistemic attitude one forms towards the conclusion.

In addition to deduction as mere reasoning, there is also a sense in which it refers to the way a belief is formed. In this sense, using deduction as a method implies that one believes the conclusion of the deductive reasoning. Deduction in this sense also implies that one believes the premises. One believes the conclusion (at least in part) because one believes the premises.<sup>11</sup> Roughly speaking, then, deduction as a belief-forming process has a tripartite structure: a condition on the input epistemic state, a condition on the reasoning based on the input epistemic states and a condition on the output epistemic state.

It is clear that extending our knowledge by competent deduction saliently targets the sense of deduction as a belief-forming mechanism, for we are interested in a belief being formed by way of inferring it from a number of premises. With this in mind, we can now ask how deductive methods should be individuated.

What I'd like to draw attention to is that the input states a deductive method accepts provides interesting options. It is clear that a deductive method which delivers beliefs as outputs should only accept beliefs as input. But as far as individuation is concerned, there is the option of going further: forming a belief based on *known premises* may

<sup>10</sup>As a case in point, Pritchard (2005: 156) casts safety not in terms of methods but in terms of ways in which a belief is formed.

<sup>11</sup>The bracketed qualification takes care of cases in which a conclusion is *epistemically overdetermined*: the subject possesses and is responsive to more than one set of sufficient reasons for the conclusion. Thanks to an anonymous referee for pointing this out.

not be the same method as forming a belief based on *believed premises*, where a believed premise will occasionally not amount to knowledge. Deductive inference from knowledge may not constitute the same method as deductive inference from belief. The point of forming a belief based on knowledge is that one transfers the good epistemic standing of the premises to the inferred conclusion. On this line of thought, inferring a conclusion from an instance of knowledge would not be the same method as inferring the same conclusion by way of the same reasoning from a belief with the same content if this belief does not constitute knowledge.

An approach along these lines interestingly differs from the approach set out in the previous section. On the present idea, the identity of the content of the input states – the premises – would not have to be held fixed. Rather, it would be the epistemic quality of the input states – knowledge versus mere belief – which would contribute to the individuation of methods.<sup>12</sup> Let us make this explicit:

**Inferential Methods (externalist version).** Deductive belief-forming methods A and B are the same only if A and B take as input premises with the same epistemic status (a method which accepts only known premises is different from a method which also accepts mere beliefs).

Applied to **Safety (method version)**, this means that if one knows the premises of one's deductive inference in the actual world, then a belief in the conclusion is safe only if the conclusion is true in all close worlds in which it is believed when inferred from known premises.<sup>13</sup> If in close worlds the premises are not known, then an inferential belief would not count as being formed by the same method if the premises are known in the actual world. It is clear that this approach blocks the counterexamples by Murphy and Alspecter-Kelly, for they require one of the premises to be false in a close world in which the same method is applied.

Before we turn to the question of whether counterexamples can be blocked in general, a word on 'competent deduction' as used in the characterization of **Closure**. It seems clear that in order to gain knowledge of a given conclusion, more is required than just getting it right. Someone who is not competent in making a certain inference and is therefore prone to get it wrong in very similar situations may not know what she inferred from something she knows. This is so even if, in fact, the conclusion follows from the premises. For this reason, one may require from competent deduction not only (i) that the conclusion in fact follows from the premises but also (ii) that in close worlds in which the subject starts a similar reasoning process from potentially different premises, the conclusion she reaches follows from the premises. Although even more may be required, for present purposes it suffices to assume that competent deduction satisfies these two conditions (see section 6 for further discussion):

**Competent Deduction.** If a subject competently deduces  $Q$  from  $P_1, \dots, P_n$ , then (i)  $Q$  follows from  $P_1, \dots, P_n$  and (ii) in all close worlds in which the subject uses the same method, the conclusion follows from the premises.

<sup>12</sup>It might be worth noting that the two approaches of individuating methods are not incompatible. In principle, methods could be individuated as finely as suggested in the previous section with an externalist constraint added on top. For the task of reconciling closure and safety, this would seem, however, unnecessarily excessive.

<sup>13</sup>As a matter of fact, it would suffice to draw a line between inference from *true belief* and inference from (any) belief. Whether individuating methods more finely in terms of known premises should be favored is something we can leave open.

Note that, strictly speaking, clause (i) is redundant. As the actual world is a close world, clause (ii) implies (i). I include it only for reasons of emphasis.

Given these considerations, it is now straightforward to see that **Safe Deduction**, as explicated in the previous section, holds.<sup>14</sup> To this end, suppose the conditions (i)–(iii) are satisfied. What we want to show is that the subject's belief in *Q* is safe in the sense of **Safety (method version)**. So, consider a close world *w* in which the subject believes *Q* by using the same method, which is, by (iii) in **Safe Deduction**, a competently executed deduction. By (i) in **Safe Deduction**, the subject knows the premises of the inference in the actual world. Therefore, by **Inferential Methods (externalist version)**, the subject knows the premises (which may differ from the actual premises) in *w* as well. Now, by **Competent Deduction**, competently using a deductive method requires that the conclusion follows from the premises in all close worlds. By assumption, that is (ii) in **Safe Deduction**, the subject competently inferred the conclusion from the premises. Therefore, the conclusion follows from the premises in *w*. Given that these premises are known, they are true and so it follows that *Q* is true in *w* as desired.

This demonstrates that individuating methods in terms of the epistemic standing of the input states is strong enough to show that beliefs competently inferred from knowledge are safe. Methods, on this picture, are thus individuated partly externally. Given that a subject can sometimes not know whether or not an input belief constitutes knowledge, they could sometimes not know which method they are actually using.

Independently of the issues surrounding deductive knowledge, it has been noticed (Broncano-Berrocal 2014; Grundmann 2018) that **Safety (method version)** is most plausible on an external individuation of methods. In this debate, where non-inferential knowledge is at issue, the external components in method individuation usually concern *instruments*, *informants* or *environmental conditions*. The externalist holds that *looking at a well-functioning clock* may be a different method than *looking at a clock (which may or may not be well-functioning)*. She also holds that *forming a belief based on a reliable informant* may be a different method than *forming a belief based on an informant (which may or may not be reliable)*. Finally, the externalist holds that *perceiving an object in epistemically friendly conditions* may be a different method than *perceiving the object under conditions which may or may not be epistemically friendly*. All this is meant to hold even if the methods are indistinguishable from the subject's internal point of view.

To illustrate the parallel between non-inferential and inferential knowledge, consider a case of a *counterfactually unreliable informant*:<sup>15</sup>

#### *Anxious informant*

It is Halloween and Jane asks John whether he will come to the party. John tells her “No”, which is the truth. *Background story*: John thought about going to the party but decided against it by flipping a coin. And Jane considered dressing up as a masked killer from the movie *Scream* but decided against it by flipping a coin. Dressing up as a masked killer would have made John very anxious. He would have been so scared that he would have answered “No” in any case when asked by the killer look-alike “Are you coming to the party?”

<sup>14</sup>As before (see fn 7), it is possible to strengthen the argument to show that competent deduction from *safe* premises yields a safe conclusion. To this end, **Safety (method version)** would have to be interpreted so that safety counts as an epistemic status to be held constant in the individuation of methods.

<sup>15</sup>The following case merges aspects of an example concerning a deceiving informant by Comesaña (2005) and a case concerning a demon overlooking an otherwise reliable clock by Kelp (2009).

It is plausible and in line with parallel cases (see Grundmann 2018: Sec. 4 for discussion) that Jane knows in the actual world that John is not coming to the party. After all, John is a reliable informant under the actual circumstances (he is not scared by Jane's actual dress). However, an externalist may disagree with Comesafía (2005) and Kelp (2009), who claim for similar cases that Jane's belief is not safe and therefore that safety is not necessary for knowledge. One might think that Jane could have easily acquired a false belief because she could have easily dressed up as a killer in which case she would have been misinformed in those close worlds in which John decides to go to the party but tells her "No" because he is scared. An externalist about method individuation like Grundmann (2018) can respond that Jane uses a different method in the problematic worlds. The method she actually uses is to be externally individuated and corresponds to something like *trusting a reliable informant*.

If one accepts the externalist's response to challenges of the present kind, then one is in a position to draw a close parallel between an externalist individuation of non-inferential methods and an externalist individuation of inferential methods. Inferring a conclusion from knowledge would be like trusting a reliable informant, while inferring a conclusion from a belief which may or may not be knowledge would be like trusting an informant who may or may not be reliable. Given this parallel, the present approach to deduction fits well with general safety-based theories of knowledge.

## 6. A residual issue

The present account is based on two main elements. External individuation of methods ensures that whether a belief is safe or not depends only on *good cases*, namely on what the subject believes in situations in which the premises are known (provided they are known in the actual world). The second element is *competent deduction*. This ensures that the subject is not prone to make a mistake in reasoning from the premises to the conclusion.

With this story in place, it can be shown that beliefs competently deduced from known premises are safe. One may wonder, however, whether this argument does not show too much. Suppose one inferred the conclusion from known premises in the actual world by an application of modus ponens. If the method itself requires the premises to be known in every world in which it is used, then it may seem that the method is *absolutely fail-proof*: as modus ponens never leads one from true premises to a false conclusion, beliefs formed by this method will be automatically true in all possible worlds. This seems too strong.<sup>16</sup>

The present issue rests on the assumption that the validity of a deductive method is an essential property of the method in question: if the conclusion is validly inferred in the actual world, only a valid method in some possible world can be identical to the method which was actually used. Admittedly, it is very tempting, from a pre-theoretical standpoint, to individuate deductive methods in this way. But I think this move should be resisted, for it does not square well with the purpose that a safety condition is supposed to play in a theory of knowledge.

Safety articulates the idea that a belief formed in a certain way could not easily have been false. Applied to a piece of deductive reasoning, this can be fleshed out by saying that the subject's way of reasoning could not easily have led her from true premises to a false conclusion. This leaves open whether the subject would have arrived at a false conclusion in a (non-close) world in which, say, her reasoning skills are seriously impaired. If this is right, a theoretically more fruitful way of thinking about deductive methods in

<sup>16</sup>I am grateful to an anonymous referee for pressing me on this.

the context of a safety-based theory of knowledge could be stated in terms of the modal properties of the subject's actual reasoning. Most prominently, this would involve the *reasoning dispositions* the subject actually has. Could the subject easily have arrived at a different, possibly false conclusion when reasoning from the actual premises? Could she easily have arrived at a false conclusion when reasoning from a similar but different premise set? On the envisaged way of individuating deductive methods, instantiating a valid inference pattern in the actual world does not by itself guarantee that one cannot go wrong in other worlds by using the same method.<sup>17</sup>

There is a limiting case which may still prove problematic. Consider logical theorems which can be validly inferred from an empty premise set. On the present notion of safety, a belief in a logical theorem (or any necessary truth for that matter) is automatically safe, for there exists no possible world in which the belief is false.

Given that safety has only been construed as a necessary condition on knowledge, one need not worry that automatic safety of logical truths would result in automatic knowledge of logical truths. However, one route for explaining why a particular belief in a logical truth does not constitute knowledge would be blocked. One could not explain the absence of logical knowledge in terms of a lack of safety.

Yet, on the picture elaborated so far, one can explain the absence of knowledge through an absence of competent deduction. If a belief in a logical truth was not competently deduced, that is if one's reasoning could easily have led one astray, then one may cite this as a reason why the subject does not know the logical truth in question.

From a slightly different approach, one could strengthen safety by requiring not only that the actual belief is true in close worlds if arrived at by the same method, but also that *similar* (Williamson 2009a) or *all* (Grundmann 2018) beliefs arrived at by the same method must be true. Then a belief in a logical theorem may not be safe because the subject could easily have arrived at a different and false belief. As far as I can see, this approach could be integrated into the present picture without jeopardizing the overall conclusion that beliefs competently deduced from knowledge are safe.

## 7. Concluding remarks

There is the plausible assumption that knowledge is closed under competent deduction in favorable circumstances. One prominent approach to knowledge takes knowledge to answer to a safety condition. Yet a number of potential counterexamples challenge this pair of assumptions. On closer inspection, it turns out that the problematic cases are best seen as challenging a safety condition on knowledge, for in these cases the subjects always seem to know the conclusion of the relevant inference.

The line of response developed in this paper is based on the observation that a plausible safety condition on knowledge must be relativized to methods for independent reasons. This provides room for maneuver: one can plausibly hold that in the problematic cases, the belief-forming methods are not held constant across possible worlds.

The difficulty with this strategy is to say something on when inferential methods count as the same. One option is to hold that the premises are essential for the way a derivative belief is formed. When understood broadly, so that even the grounds of one's direct premises must be kept constant, this way of individuating methods can be used to defend closure under competent deduction. However, one faces the objection

<sup>17</sup>If a valid deduction is carried out *competently*, then it follows that the subject's reasoning is sound in close worlds (see condition (ii) in **Competent Deduction**). What does not follow, however, is that the subject does not make a mistake in a non-close world.

that this proposal individuates methods too finely by restricting the range of input states that a method accepts to only one kind.

A better way of individuating inferential methods is to partly individuate them externally. It would be partly constitutive of the way a derivative belief is formed that it was deduced from something the subject knows. Deduction from knowledge is not the same method as deduction from belief. On this account, the premises of an inference are seen in close analogy to the environment in which perception takes place. Just as one does not always know whether the environment one is in is epistemically friendly or not, so one will not always be able to know whether what one treats as premises is of the good (knowledge) or the bad (mere belief) kind.

This account of deduction rebuts the counterexamples and provides the means for a general argument that beliefs competently deduced from knowledge are always safe. It also reveals that deduction faces the same kind of intricate issues of individuation as non-inferential belief-forming methods such as perception. Thus, deduction does not come easy: explaining deductive knowledge is (almost) as hard as explaining perceptual knowledge.<sup>18</sup>

## References

- Alspector-Kelly M.** (2011). 'Why Safety Doesn't Save Closure.' *Synthese* **183**, 127–42.
- Bernecker S.** (2012). 'Sensitivity, Safety, and Closure.' *Acta Analytica* **27**, 367–81.
- Broncano-Berrocal F.** (2014). 'Is Safety in Danger?' *Philosophia* **42**, 63–81.
- Comesaña J.** (2005). 'Unsafe Knowledge.' *Synthese* **146**, 395–404.
- Dretske F.** (1970). 'Epistemic Operators.' *Journal of Philosophy* **67**, 1007–23.
- Dretske F.** (2005). 'The Case against Closure.' In M. Steup and E. Sosa (eds), *Contemporary Debates in Epistemology*, pp. 13–26. Oxford: Blackwell.
- Goldman A.I.** (1976). 'Discrimination and Perceptual Knowledge.' *Journal of Philosophy* **73**, 771–91.
- Grundmann T.** (2018). 'Saving Safety from Counterexamples.' *Synthese*, published as online first.
- Hawthorne J.** (2005). 'The Case for Closure.' In M. Steup and E. Sosa (eds), *Contemporary Debates in Epistemology*, pp. 26–43. Oxford: Blackwell.
- Hawthorne J. and M. Lasonen-Aarnio** (2009). 'Knowledge and Objective Chance.' In P. Greenough and D. Pritchard (eds), *Williamson on Knowledge*, pp. 92–108. Oxford: Oxford University Press.
- Kelp C.** (2009). 'Knowledge and Safety.' *Journal of Philosophical Research* **34**, 21–31.
- Kvanvig J.L.** (2004). 'Nozickian Epistemology and the Value of Knowledge.' *Philosophical Issues, Epistemology* **14**, 201–18.
- Kvanvig J.L.** (2006). 'Closure Principles.' *Philosophy Compass* **1**, 256–67.
- Murphy P.** (2005). 'Closure Failures for Safety.' *Philosophia* **33**, 331–34.
- Murphy P.** (2006). 'A Strategy for Assessing Closure.' *Erkenntnis* **65**, 365–83.
- Nozick R.** (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Pritchard D.** (2005). *Epistemic Luck*. New York, NY: Oxford University Press.
- Pritchard D.** (2014). 'The Modal Account of Luck.' *Metaphilosophy* **45**, 594–619.
- Sosa E.** (1999). 'How to Defeat Opposition to Moore.' *Noûs* **33**, 141–53.
- Williamson T.** (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson T.** (2009a). 'Probability and Danger.' *The Amherst Lecture in Philosophy* **4**, 1–35. <http://www.amherstlecture.org/williamson2009/>.
- Williamson T.** (2009b). 'Replies to Critics.' In P. Greenough and D. Pritchard (eds), *Williamson on Knowledge*, pp. 279–384. Oxford: Oxford University Press.

<sup>18</sup>I would like to thank Roman Heil, Jakob Koscholke, Patricia Rich, Sergui Spatan and Jacques Vollet for many helpful comments. Thanks are also due to an anonymous referee, whose careful comments led to significant improvements of the paper. Research for this paper profited from the generous support of the *Deutsche Forschungsgemeinschaft* (SCHU 3080/3-1/2) and was conducted within the Emmy-Noether research group 'Knowledge and Decision'.

MORITZ SCHULZ is Junior Professor of Theoretical Philosophy at the University of Hamburg, Germany. His research includes the philosophy of language and logic, epistemology as well as decision theory. He currently conducts a research group entitled 'Knowledge and Decision', which studies the role of knowledge in (rational) decision-making. His book, *Counterfactuals and Probability*, came out with Oxford University Press in 2017.