Sensitivity, Safety, and Epistemic Closure

Bin Zhao

Abstract It has been argued that an advantage of the safety account over the sensitivity account is that the safety account preserves epistemic closure; while the sensitivity account implies epistemic closure failure. However, the argument fails to take the method-relativity of the modal conditions on knowledge, viz, sensitivity and safety, into account. In this paper, I argue that the sensitivity account and the safety account are on a par with respect to epistemic closure once the method-relativity of the modal conditions is taken into account. Therefore, epistemic closure is no longer an arbiter in the debate.

Keywords Belief-formation methods · Competent deduction · Epistemic closure · Inferential beliefs · Knowledge · Safety · Sensitivity · The generality problem

1. Introduction

Modal epistemologies maintain that there must be a modal connection between one’s belief and the truth for the belief to count as knowledge: one not only should hold a true belief in the actual world, but also should avoid holding a false belief across a specified range of nonactual possible worlds. Among the main modal epistemologies in the literature are the
sensitivity account of knowledge\(^1\) and the safety account of knowledge\(^2\).

Though both accounts handle a wide range of cases involving knowledge-precluding epistemic luck as well as cases of knowledge,\(^3\) it has been argued that an advantage of the safety account over the sensitivity account is that the safety account preserves epistemic closure; while the sensitivity account implies epistemic closure failure. In this paper, I demonstrate that such an argument fails to take into account that the modal conditions are usually relativized to belief-formation methods. Once we take the method-relativity of the modal conditions into account, it should be clear to us that the sensitivity account and the safety account are on a par with respect to epistemic closure, viz., either both accounts preserve epistemic closure or both accounts imply epistemic closure failure. Therefore, epistemic closure cannot be used to adjudicate between the sensitivity and the safety accounts of knowledge.\(^4\)


\(^3\) Nonetheless, it has been argued that the sensitivity account, unlike the safety account, fails to account for inductive knowledge, See Pritchard (2008b) and Sosa (1999b). For the argument that the sensitivity account is able to account for inductive knowledge, see Wallbridge (2018). For the argument that the safety account also fails to account for inductive knowledge, see Zhao & Baumann (2021).

2. Sensitivity and Safety

Consider a scenario where one looks at a reliable clock in normal lighting conditions and thus forms a true belief that it is now 12:00. Since the clock is reliable and the lighting conditions are normal, the belief counts as knowledge. Consider another scenario where one looks at a clock that stops 12 hours earlier in normal lighting conditions and thus forms a true belief that it is now 12:00. The belief is true as a matter of luck and thus does not count as knowledge.

The lesson is that luckily true beliefs do not count as knowledge. To eliminate luckily true beliefs from the realm of knowledge, an anti-luck condition on knowledge is needed. A natural idea is that if one’s belief is true as a matter of luck, then one would still believe the same thing even if it turns out to be false. As Murphy and Black write,

“How, then, do we keep luckily acquired beliefs from counting as knowledge? We must demand more of S than that she respond appropriately to her environment by suitably forming the true belief that P. One idea is to demand that S respond appropriately in environments where it is not the case that P” (Murphy & Black, 2012, p. 30).

This idea motivates the sensitivity account of knowledge. According to this account, S knows that p only if S’s belief in p is sensitive, that is, S would not believe that p if p were false, or formally ¬p ⊨ ¬Bp (“⊨” denotes the subjunctive conditional connective) (Nozick, 1981, p. 177).

The sensitivity condition is rendered as follows under the Lewis-Stalnaker analysis of

---

subjunctive conditionals (Lewis, 1973a, 1973b; Stalnaker, 1968),

**SENSITIVITY**: S’s belief in \( p \) is sensitive just in case, in the closest possible world where \( p \) is false, S does not believe that \( p \).

This makes us consider whether \( S \) holds a false belief in \( p \) in the closest possible world where \( p \) is false. If \( S \) believes that \( p \) in the possible world, then \( S \)’s belief in \( p \) is insensitive. Thus, the belief is luckily true and \( S \) does not know that \( p \). If \( S \) does not believe that \( p \) in the possible world, then \( S \)’s belief in \( p \) is sensitive. Thus, the belief is non-luckily true and \( S \) knows that \( p \) unless it exhibits some non-modal shortcomings that would deprive it of the status of knowledge.

Another idea is that if one’s belief is true as a matter of luck, then the belief could easily have been false. This idea motivates the safety account of knowledge. According to this account, \( S \) knows that \( p \) only if \( S \)’s belief in \( p \) is safe, that is, only if \( S \) could not easily have falsely believed \( p \), or formally \( \Box p \Rightarrow p \).

The safety condition is usually rendered as follows,

**SAFETY**: \( S \)’s belief in \( p \) is safe just in case, in all nearby possible worlds where \( S \) forms a belief that \( p, p \) is true.

This makes us consider whether \( p \) is true in nearby possible worlds where \( S \) believes that \( p \). If \( p \) is false in some of these possible worlds, then \( S \)’s belief in \( p \) is not safe. Thus, the belief is luckily true and \( S \) does not know that \( p \). If \( p \) is true in all these possible worlds, then \( S \)’s belief

---

in \( p \) is safe. Thus, the belief is non-luckily true and \( S \) knows that \( p \) unless it exhibits some non-modal shortcomings that would deprive it of the status of knowledge.

Both SENSITIVITY and SAFETY produce a welcome result in the above scenarios. For SENSITIVITY, one knows the correct time in the first scenario because one’s belief is sensitive. In the closest possible world where it is not 12:00, i.e., the possible world where one looks at the clock one minute earlier or later, the reliable clock would not indicate 12:00, thus one does not believe that \( \text{it is now 12:00} \) by reading the reliable clock. One does not know the correct time in the second scenario because one’s belief is insensitive which accounts for why it is true as a matter of luck. In the closest possible worlds where it is not 12:00, i.e., the possible world where one looks at the clock one minute earlier or later, the stopped clock still indicates 12:00, thus one believes that \( \text{it is now 12:00} \) by reading the stopped clock. For SAFETY, one knows the correct time in the first scenario because one’s belief is safe. In all nearby possible worlds where one believes that \( \text{it is now 12:00} \) on the clock, it is 12:00. One does not know the correct time in the second scenario because one’s belief is unsafe. There are some nearby possible worlds where he looks at the clock one minute earlier and believes that \( \text{it is now 12:00} \) on the clock, while the time is 11:59. In a word, both accounts handle cases involving knowledge-precluding epistemic luck as well as cases of knowledge nicely.

However, both SENSITIVITY and SAFETY run into counterexamples quickly. For instance,

DACHSHUND: “I look at a nearby dachshund and truly believe that what I see is a dog.

Had I not been seeing a dachshund, I would have been seeing a wolf, and would have
falsely believed myself to be seeing a dog” (Goldman, 2009, p. 80).⁷

Intuitively, I know \textit{what I see is a dog} despite my falsely believing so in the closest possible world where what I see is not a dog, i.e., what sees in that possible world is a wolf. In addition, there is a nearby possible world where I falsely believe that \textit{what I see is a dog}. Thus, we have a counterexample for both SENSITIVITY and SAFETY.

Once it is noted that the wolf does not look like the dog, we may want to relativize the sensitivity condition to the belief-formation method to get rid of the counterexample. If the condition is relativized in that way, then a possible world should be taken into account only if it is a possible world where \(S\) uses the same method as that in the actual world; while possible worlds where \(S\) uses a different method are irrelevant. That is to say, which possible worlds are relevant depends on the individuation of belief-formation methods. In DACHSHUND, the belief-formation methods are individuated by reference to perceptual appearances. Since the wolf appears to be different from the dog, the subject’s belief in the counterfactual case is formed on a different method from that in the actual case (Goldman, 2009, p. 81). Therefore, the subject’s belief is sensitive if the sensitivity condition is relativized to the belief-formation methods.⁸ The sensitivity condition is thus rendered as follows,

\textbf{SENSITIVITY}^M: \(S\)'s belief in \(p\) which is formed on method \(M\) is sensitive just in case, in the closest possible world where \(p\) is false and \(S\) uses \(M\) to form a belief whether or not

---

⁷ Similar counterexamples include Alfano’s (2009) REDWOOD, Goldman’s (1976) JUDY & TRUDY, and Nozick’s (1981) GRADNMOTHER.
⁸ In addition to discharging putative counterexamples such as DACHSHUND, there are other motivations for relativizing the sensitivity condition to belief-formation methods. For instance, there is criticism of the sensitivity account that it cannot account for higher-order knowledge, see DeRose (1995), Huemer (2001), Kripke (2011), Melchior (2015, 2017b), Sosa (1996, 1999b, 2002), Vogel (1987, 2000, 2007, 2012), Williamson (2000), and Zalabardo (2012). Nonetheless, Bjerring & Gundersen (2020) and Wallbridge (2017, 2018a) argue that the criticism fails to take the method-relativity of the sensitivity condition into account.
Similarly, the safety is usually relativized to belief-formation methods to avoid counterexamples such as DACHSHUND. The safety condition is thus rendered as follows,

\[ \text{SAFETY}^M: S's \text{ belief that } p, \text{ formed on method } M, \text{ is safe, if and only if, in all nearby possible worlds where } S \text{ forms a belief that } p \text{ on method } M, p \text{ is true.} \]

If the safety condition is thus relativized to belief-formation methods, then counterexamples such as DACHSHUND would be discharged similarly.

3. The Consideration of Epistemic Closure

Though both accounts handle a wide range of cases involving knowledge-precluding epistemic luck as well as cases of knowledge, it has been argued that the sensitivity account implies epistemic closure failure (Kripke, 2011; Luper, 2012; Pritchard 2002, 2005, 2008b; Sosa, 1999b, 2004; Vogel, 1987; Williamson, 2000); while the safety account preserves epistemic closure (Luper, 2012; Pritchard 2002, 2005, 2008b; Sosa, 1999b; 2004). Thus, the consideration of epistemic closure is in favor of the safety account.\(^9\)

Epistemic closure is the idea that knowledge is closed under competent deduction in the sense that we can extend our knowledge by competently deducing consequences from what is already known. For instance, if I know that I am looking at a red barn, then I can know that I am looking at a barn by deduction. To put it formally,

Epistemic Closure: If S knows p and competently deduces q from p, thereby coming to

---

\(^9\) For dissenters who think that the safety account also implies epistemic closure failure, see Alspector-Kelly (2011), Kvanvig (2008b), and Murphy (2005, 2006).
believe \( q \), while retaining \( S \)’s knowledge that \( p \), \( S \) comes to know that \( q \) (Hawthorne, 2005, p. 29).\(^{10}\)

This principle seems to be a plain truth to the effect that many epistemologists take it for granted without further arguments. For instance, Steven Luper writes that “virtually everyone agrees that the closure principle, suitably formulated, is highly plausible” (Luper, 2012, p. 222). After all, what can be plainer than knowing that \textit{I am looking at a barn} by deducing it from my knowledge that \textit{I am looking at a red barn}?

Nonetheless, it has been argued that \( S \) can have an insensitive belief in \( q \), which is competently deduced from \( p \) which \( S \) knows to be true. For instance, I know that \textit{I have hands} because my belief in it is sensitive, viz., in the closest possible world where I do not have hands, e.g., a possible world where I lost my hands in a severe car accident, I do not believe that \textit{I have hands}. However, I do not know that \textit{I’m not a handless brain-in-a-vat} because my belief in it is insensitive, viz., in the closest possible world where I’m a handless brain-in-a-vat, I still believe that \textit{I’m not a handless brain-in-a-vat}, though \textit{I’m not a handless brain-in-a-vat} is competently deduced from \textit{I have hands} which I know to be true. Therefore, the sensitivity account implies epistemic closure failure (Pritchard 2002, 2005, 2008b; Sosa 1999b; Vogel, 1987; Williamson, 2000).

In turn, let us see how the safety account performs when it comes to epistemic closure. According to the safety account, I know that \textit{I have hands} because my belief in it is safe, viz., in nearby possible worlds where I believe that \textit{I have hands}, it is true that I have hands.

\(^{10}\) This formulation needs some further qualifications. For example, \( S \) learns of no undefeated defeater for \( q \). For the sake of simplicity, I shall avoid further complications here. For related discussions, see David and Warfield (2008), Kvanvig (2006), Luper (2020), Warfield (2004).
because possible worlds where I falsely believe so, e.g., the possible worlds where I’m a handless brain-in-a-vat, are very remote. I also know that I’m not a handless brain-in-a-vat because my belief in it is safe, viz., in nearby possible worlds where I believe that I’m not a handless brain-in-a-vat, it is true that I’m not a handless brain-in-a-vat given that the possible worlds where I’m a handless brain-in-a-vat are very remote. Thus, it has been argued that the safety account is better placed than the sensitivity account to preserve epistemic closure which, in turn, is one of the main reasons in favor of the safety account.

4. The Method-Relativity of Modal Conditions and Epistemic Closure

However, the aforementioned argument does not pay enough attention to the method-relativity of modal conditions. In a case where S competently deduces q, e.g., I’m not a handless brain-in-a-vat, from p, e.g., I have hands, which S knows to be true. S not merely believes that q, but believes it on some belief-formation method. When evaluating whether the belief that q is sensitive or not, the argument examines the closest possible world where the target proposition is false, e.g., the closest possible world where I’m a handless brain-in-a-vat. Nonetheless, what we should examine is the closest possible world where the target proposition is false and the subject uses the same method as that in the actual world. It is unclear whether S’s belief that q is sensitive or not once the method-relativity of the modal conditions is taken into account.

There are different ways to individuate the belief-formation method: it could be individuated as inference from the belief that p (Alspector-Kelly, 2011; Nozick, 1981), or individuated as inference from the justified belief that p, or individuated as inference from the
true belief that \( p \); or individuated as inference from knowledge that \( p \) (Hirvelä, 2020; Schulz, 2020; Williams, 1991), etc.\(^{11}\) Depending on whether the truth of \( p \) is ensured by the individuation, the individuations could be classified into two sorts: the individuations that do not ensure the truth of \( p \), e.g., inference from the belief that \( p \) or inference from the justified belief that \( p \). If the method is thus individuated, then one is able to use the method no matter whether \( p \) is true or not; and the individuations that ensure the truth of \( p \), e.g., inference from the true belief that \( p \) or inference from knowledge that \( p \). If the method is thus individuated, then one is able to use the method only if \( p \) is true.

Thus, which possible world is relevant when evaluating whether the target belief is sensitive or not depends on how we individuate the belief-formation method of the target belief. How do we determine which individuation is the correct one? We seem to encounter the generality problem here. Instead of committing myself to a specific individuation of the belief-formation methods, I shall argue that the sensitivity account and the safety are on a par with respect to epistemic closure no matter whether the belief-formation method is individuated such that the truth of \( p \) is ensured or not.

4.1 Individuations that Do Not Ensure the Truth of \( p \)

In a case where \( S \) competently deduces \( q \) from \( p \) which \( S \) knows to be true, if the belief-formation method is individuated such that the truth of \( p \) is not ensured, e.g., inference from the belief that \( p \) or inference from the justified belief that \( p \), then it is the case that \( S \)’s belief that \( q \) could be both insensitive and unsafe.

For the belief to be insensitive, $S$ needs to believe that $q$ in the closest possible world where $q$ is false and $S$ uses the same method as that in the actual world, e.g., inference from the belief that $p$ or inference from the justified belief that $p$, to form a belief whether or not $q$. Since $p$ entails $q$, $p$ is also false in that possible world. In addition, $S$ should believe that $p$ in that possible world. Otherwise, $S$ can not infer $q$ from $p$. Thus, $S$ believes that $p$ in a possible world where $p$ is false. Since $S$ knows that $p$, $S$’s belief that $p$ in that possible world should be formed on a different method from that in the actual world or that possible world is not the closest possible world where $p$ is false and $S$ uses the same method as that in the actual world. As long as this desideratum is satisfied, we would have an epistemic closure failure case for the sensitivity theorists.

For the belief that $q$ to be unsafe, $q$ needs to be false in some nearby possible world where $S$ believes that $q$ on the same method as that in the actual world, e.g., inference from the belief that $p$ or inference from the justified belief that $p$. Since $p$ entails $q$, $p$ is also false in that possible world. In addition, $S$ should believe that $p$ in that possible world. Otherwise, $S$ can not infer $q$ from $p$. Thus, $p$ is false in some nearby possible world where $S$ believes that $p$. Since $S$ knows that $p$, $S$’s belief that $p$ in that possible world should be formed on a different method from that in the actual world. As long as this desideratum is satisfied, we will have an epistemic closure failure case for the safety theorists.

A revised version of Goldman’s DACHSHUND satisfies the aforementioned desideratum,

**DACHSHUND II**: I look at a nearby dachshund and truly believe that *what I see is a dog*. Had I not been seeing a dachshund, I would have been seeing a wolf, and would
have falsely believed myself to be seeing a dog. I then infer that what I see is not a wolf.\footnote{Similar examples include Alspector-Kelly’s (2011) LAMBORGHINI and Kripke’s (2011) RED BARN.}

Intuitively, I know \textit{what I see is a dog}. What about my inferential belief that \textit{what I see is a wolf}? In the closest possible world where \textit{what I see is a wolf} and I use the same method as that in the actual world, e.g., inference from the belief that \textit{what I see is a dog} or inference from the justified belief that \textit{what I see is a dog}, to form a belief whether what I see is a wolf or not, I believe \textit{what I see is not a wolf}. Thus, my belief that \textit{what I see is not a wolf} is insensitive. In addition, in some nearby possible world where I believe \textit{what I see is not a wolf} on the same method as that in the actual world, e.g., inference from the belief that \textit{what I see is a dog} or inference from the justified belief that \textit{what I see is a dog}, what I see is a wolf. Thus, my belief that \textit{what I see is not a wolf} is unsafe. Therefore, my belief that \textit{what I see is not a wolf} is both insensitive and unsafe and thus falls short of knowledge though it is competently deduced from what I know to be true.

Here is a recipe to construct an example that shows that the sensitivity account and the safety account imply epistemic closure failure: First, construct a case where $S$ competently deduces $q$ from $p$ which $S$ knows to be true on some belief-formation method. Second, make sure that there is a possible world where $S$ falsely believes that $p$ on a different method. Third, make sure that $S$ falsely believes that $q$ by competent deduction from $S$’s belief that $p$ in that possible world. Forth, make sure that the possible world is a nearby one and it is also the closest possible world where $q$ is false. Once these conditions are satisfied, we will have an example showing that both accounts imply epistemic closure failure.
In a word, both accounts imply epistemic closure failure if the belief-formation method is individuated such that the truth of $p$ is not ensured.

4.2 Individuations that Ensures the Truth of $p$

In a case where $S$ competently deduces $q$ from $p$ which $S$ knows to be true, if the belief-formation method is individuated such that the truth of $p$ is ensured, e.g., inference from the true belief that $p$ or inference from knowledge that $p$, then it is the case that $S$’s belief that $q$ is both insensitive and unsafe.

If the belief-formation method is thus individuated, then $p$ is true in all possible worlds where $S$ forms a belief that $q$ on the same method as that in the actual world. Since $p$ logically entails $q$, $q$ is also true in these possible worlds. Thus, $S$’s belief that $q$ is both sensitive and safe because there is no possible world where $S$ falsely believes that $q$ on the same method as that in the actual world.

In a word, both accounts preserve epistemic closure if the belief-formation method is individuated such that the truth of $p$ is ensured.

5. Conclusion

Whether the sensitivity account and the safety account preserve epistemic closure or not depends on how the belief-formation method of inferential beliefs is individuated. In a case where $S$ competently deduces $q$ from $p$ which he knows to be true. If the method is individuated such that the truth of $p$ is ensured, e.g., inference from the true belief that $p$ or inference from knowledge that $p$, then $S$’s belief that $q$ is both sensitive and safe; if the
method is individuated such that the truth of p is not ensured, e.g., inference from the belief that p or inference from the justified belief that p, then S’s belief that q could be both insensitive and unsafe. The conclusion is that either both accounts preserve epistemic closure or both accounts imply epistemic closure failure. Therefore, the consideration of epistemic closure does not help us to advance the debate in modal epistemology.13

REFERENCES


13 I would like to thank Sven Bernecker, Duncan Pritchard, and an anonymous referee for helpful comments and suggestions.


——. (2008). “Defending a Sensitive Neo-Moorean Invariantism.” In Vincent F. Hendricks...


1–52.


Goldberg, Sanford. (2012). “Sensitivity from Others.” In Kelly Becker & Tim Black (eds.),  
*The Sensitivity Principle in Epistemology*. New York: Cambridge University Press,  
43–65.


Sensitivity Principle in Epistemology*. Cambridge: Cambridge University Press,  
193–206.

Grundmann, Thomas. (2020). “Saving Safety From Counterexamples.” *Synthese*, 197:


